AD-A042 335 NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
SURVEY OF TECHNICAL MANUAL READABILITY AND COMPREHENSIBILITY, (U)
JUN 77 T E CURRAN

UNCLASSIFIED NPRDC-TR-77-37 NL

1 OF 2

AD
A042335

ADA042335

(11) NW

NPRDC TR 77-37

JUNE 1977

SURVEY OF TECHNICAL MANUAL READABILITY AND COMPREHENSIBILITY

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

NPRDC TR 77-37

June 1977

SURVEY OF TECHNICAL MANUAL READABILITY AND COMPREHENSIBILITY

Thomas E. Curran

Reviewed by
Robert E. Blanchard
Edwin G. Aiken

Approved by
James J. Regan
Technical Director

Navy Personnel Research and Development Center
San Diego, California 92152

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPRDC TR 77-37	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) SURVEY OF TECHNICAL MANUAL READABILITY AND COMPREHENSIBILITY.		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Thomas E. Curran		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N00167-76-WR-7-0104
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		12. REPORT DATE (11) June 1977
		13. NUMBER OF PAGES 150
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (12) 148p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(14) NPRDC-TR-77-37

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Readability
Comprehensibility
Comprehension
Technical Manuals
Graphics
Maintenance

(Many

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

There is increasing concern that Navy personnel have severe difficulty in reading and comprehending the training and technical manuals required on their jobs. Various reports have indicated that the Reading Grade Level of many Navy manuals range to the twelfth grade and above. Other reports indicate that the reading ability of the average person entering the naval service is at approximately the tenth grade level. In order to attempt to correct this mismatch, a large-scale program, originally termed the Navy Technical Manual

but

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

iii 390 772

cont fr p iii

System ~~(NTMS)~~ and recently renamed the Navy Technical Information Presentation Program, ~~(NTIPP)~~ came into being. The goal of this program is goal to investigate problems associated with technical manuals and to recommend actions to alleviate ~~these problems~~ them. This report concerns with ~~one aspect of this problem~~, the reading difficulty of manuals and how it might be improved. Its focus is on an extensive survey of the research literature, theoretical discussions, and existing state-of-the-art techniques and devices as they relate to the prediction (i.e., assessment) of the difficulty of manuals and the production (i.e., writing and printing) of manuals at difficulty levels appropriate to the intended user. Both the comprehensibility of text and that of graphics were addressed. Recommendations are made, based on the findings of this survey are put forth for consideration.

FOREWORD

This work unit was performed in support of the Navy Technical Manual System (NTMS) (now titled the Navy Technical Information Presentation Program, NTIPP) under the auspices of the Naval Ship Research and Development Center, Bethesda, Maryland. The goal of this program is to develop a system of procedures and equipments designed to support and improve the utility, preparation, revision, storage, distribution, and overall management of technical data for the mid-1980 time period. A major problem in the overall technical documentation system is the determination of the reading difficulty level of manuals and the development of procedures for matching that difficulty level to the abilities of the intended users of the manuals. Navy Personnel Research and Development Center, San Diego, was tasked with investigating this problem. This study focuses on identifying those characteristics of both text and graphics which have causal relationships to on-the-job comprehension of technical manuals. It is hoped that this work contributes to the ultimate success of the NTIP Program.

Appreciation is expressed to Dr. George R. Klare of Ohio University for his contribution of the annotated bibliography dealing with the parameters of writing as they affect comprehension. In addition, the comments and recommendations of Dr. Klare on all facets of the report were very welcome and greatly appreciated.

J. J. CLARKIN
Commanding Officer

SUMMARY

Problem

Navy personnel have expressed continuing dissatisfaction with the quality of training manuals and technical manuals (TMs) required to perform their duties. This discontent has involved virtually every facet of the TM cycle: untimely receipt, lack of necessary technical information, difficulty level of the content, etc. A major aspect of this problem is that many of the Navy's manuals are written at a "reading grade level (RGL)" considerably above the reading ability of the personnel who must use such manuals. It is therefore vital that research be undertaken to determine ways in which the writers of Navy manuals can direct the level of difficulty of their work to the level of ability of its intended user.

The volume of technical manual materials in current use by the Navy is estimated to be in excess of 70,000,000 pages. Training manuals other than TMs may easily equal or surpass this amount. Using this vast volume of materials are approximately 1,000,000 personnel, with reading abilities extending from perhaps the fourth or fifth grade level to that of a graduate engineer. One can easily envision the turmoil when a TM targeted for the former reading level finds its way to the latter--or more critically, when the reverse is true. The difficulty level of technical documentation--"readability" in its most general sense--is a major determining factor as to whether or not technicians will use (and persist in the use of) TMs. Steps must be taken to ensure that the difficulty level of both text and graphics is matched to the reading abilities of the user population.

Approach

Guiding the overall approach to the problem were a set of "themes" coming from a recent conference on "Improved Information Aids for Technicians" in Arlington, Virginia. The essential thrust of this conference was Job Performance Aids (JPAs), of which maintenance manuals are but one type. The most pertinent of these themes were as follows:

1. An overwhelming need exists for a "model" or "structure" of the maintenance problem in general. Piecemeal efforts will only hinder the attainment of our overall objectives.

2. Research in many aspects of maintenance documentation is vitally needed, but it is widely known that the application of research findings lags 5, 10, or more years behind the work itself. We <u>must</u> use today's technology for tomorrow's needs and accelerate the implementation of tomorrow's research into near future application.

3. A "head-book tradeoff" in terms of cost-effectiveness must be available. Neither all training and no documentation nor its reciprocal are satisfactory. Again, no <u>model</u> exists to guide this tradeoff.

4. Users must acquire maintenance documentation in a more timely manner (which virtually all agree is a thorny problem).

5. The tendency to think in terms of traditional manuals versus fully proceduralized JPAs versus other new concepts in information transmission must be overcome, and the best features of each must be used in various combinations. The need for standardization must be kept in mind even though greater variability may result from this concept, at least at the outset.

The approach, using these themes for guidance, took two main thrusts. First, a comprehensive review of the literature was made in the area, including books, journal articles, technical reports, and unpublished manuscripts. Selected annotated bibliographies of these sources are included in the report. Secondly, personal contact was made with a number of companies and individuals who are working or have worked in this field to solicit their ideas on all aspects of the problem.

Conclusions and Recommendations

1. There is little empirical evidence regarding those parameters of textual material which have a causal relationship to comprehension.
RECOMMENDED: Instigate research and development to identify those parameters of technical writing which affect comprehension.

2. It is of major importance that we obtain maximum use of existing technology in the technical writing process while pursuing research and development aimed at improving the process.
RECOMMENDED: Continue test and evaluation of the Navy TRUMP system and increase its capabilities in the area of assessment and improvement of readability and comprehensibility.

3. Technical writers require a simple but effective means of determining the difficulty of their writing at an early stage, and a much more sensitive index is required at later stages of processing, such as editing.
RECOMMENDED: Identify and implement a simple readability formula for use by writers and develop the most sensitive possible technique for assessing readability and comprehensibility at later draft stages.

4. Technical writers and illustrators require, in addition to assessment devices, guidelines to assist them in meeting the reading difficulty specifications of their work while maintaining high comprehensibility.
RECOMMENDED: Using empirical data, construct a writer's guide or "job performance aid" for technical writers that is both usable by them and acceptable to them.

5. The large volume of technical materials with which we must deal argues for a computerized feedback system for technical writers.
RECOMMENDED: Develop and implement a cost-effective computerized feedback system which will provide readability measurement as well as such data as word frequency counts and alternative terminology.

6. Technical manual production is very much a "piecemeal" effort with insulation between writer and engineer, writer and illustrator, writer and editor, etc.
RECOMMENDED: Develop and implement an integrated technical manual production system with interfaces at each step from design engineering, through first draft, to the user.

7. Virtually nothing is known regarding the parameters of various types of technical graphics which directly affect comprehension.
RECOMMENDED: Institute basic research to determine the parameters of technical graphics which have causal relationships to comprehension.

8. Technical manual graphics, despite their volume and the computer state-of-the-art, require a great deal of human intervention in their production.
RECOMMENDED: Develop automated procedures for the production of technical graphics to the fullest possible extent, or identify such procedures if they currently exist.

CONTENTS

TABLE

LIST OF FIGURES

INTRODUCTION

Background

Comprehensibility of Written Materials: A Brief History

Communicating comprehensibly has no doubt been a concern since symbols were first used to transmit information. Indeed, the very effort of inventing symbols intrinsically reflects this concern. A person who commits to paper (or other medium) any potentially meaningful information does so with the obvious intent of communicating that information to other persons--probably a limited and select audience. To make his efforts readable and comprehensible would seem intuitively to be a willful effort.

Some of these efforts have been notably successful; others, as the reader will see, have been dismal failures. In either case, the "audience" is a significant factor. Author James A. Maxwell states the situation in the following way:

> Man's strong urge to communicate with his fellow humans has always been complicated by his equally powerful compulsion to limit comprehension to a particular group. Thus, as he develops the printing press and Esperanto to further understanding between peoples, he works equally hard on cryptology and plays like "Waiting for Godot," to confuse all but the chosen few. (1960, p. 33)

In any event, there is a continuum throughout history from the most highly successful communication to that which even today baffles our sophisticated society.

Prehistoric man, in drawing crude symbols on the walls of his cave, might be considered to have been extremely successful in communicating the location of water, the abundance of game, or the existence of a hostile neighboring tribe. If this communication was as effective as it might appear, it was probably due to the relatively small number of symbols involved. One would expect, further, that knowledge of the meaning of each symbol was restricted to a rather close-knit group, and that such meanings were established by word of mouth.

Hamilton (1970) points out that in ancient Syria, many centuries before Christ, ". . . men of intelligence and sophistication rolled and stamped out images on clay." He goes on to say, that, even after close inspection, ". . . it was obvious that only a small number of experts could have the patience and perception to interpret and appreciate the mighty messages of these tiny symbols." Yet, they were meaningful to those of that day, and, with perseverance, are now meaningful to modern man. As Hamilton puts it, "In their hard and durably etched surfaces was preserved a greatly condensed record of powerful traders, ruling families, and wise oracles who framed the rules under which societies flourished and declined."

A more familiar example of such effective communication is cryptology. Codes of various types have been extremely effective in communicating with a selected audience, particularly since the advent of electronic devices but probably dating back to the earliest forms of warfare. Witness, for example, the American Indian's use of smoke signals, a code which was virtually meaningless to most white men, but which rather precisely guided the behavior of geographically separated bands of Indians. The Japanese codes being used prior to the attack on Pearl Harbor were quite effective, although in time they were "broken." It is of interest to note, however, that, even when the code itself was understood, the meaning of the underlying communication was not easily comprehensible by American cryptologists--persons who were not members of the intended audience.

In a lighter vein, there is an example of communication arising from a fading facet of American folklore, which was apparently both simple and quite effective. This is "hobo symbology," by which itinerants communicated characteristics of their environment to one another. In discussing this "language," Maxwell (1960) points out that few "ordinary citizens" would give a second glance to a wooden fence with some vague symbol chalked upon it. To the hobo, however, these symbols communicated information that was important, or even vital, to his lifestyle. For example if he saw the symbol in Figure 1a, which means "an officer of the law lives here," or that in Figure 1b, which means "an open eye--authorities alert here," he would no doubt go swiftly on his way. However, he would be happy to see the symbol in Figure 1c, for this tells him that the nearby house is "OK for a handout." But, as with any mode of communication, mixed results may be obtained by combining symbols. The symbol in Figure 1d, which means "a stack of logs," together with that in Figure 1c would tell a hobo that a meal was in store, but that he would have to work for it.

Thus far, we have been considering written communication composed of relatively few discrete symbols, which are quite different from what we call "words." When man began to combine discrete symbols into words in order to convey more complex meanings, the problems of readability and comprehensibility became more acute. With the development of the printing press beginning in approximately 1550, these conditions were even more aggravated, because the intended audience grew along with the technology. As long as the printed word was laboriously penned onto a document, which often would be one-of-a-kind, the audience remained limited, and writing was virtually a matter of equals "speaking" to equals. However, with the advent of the printing press, a rather unique event occurred: literacy was at once both the cause and the effect of mass communication. The man in the street was curious about the world around him beyond his everyday domain, and printed matter helped to satisfy that curiosity. Printed matter became available, and the general public was anxious to make use of it.

Perhaps the best example of written communication coming into widespread usage is the Bible--the number one "best-seller" of all time. Men had listened for hundreds of years to the words of the Bible before ever seeing, or at least being able to read, those words in print. While "readability" does not apply to the spoken word, comprehensibility certainly does. And, even within the Bible, words of caution were spoken in an attempt to achieve it. St. Paul, in speaking to the Corinthians, said:

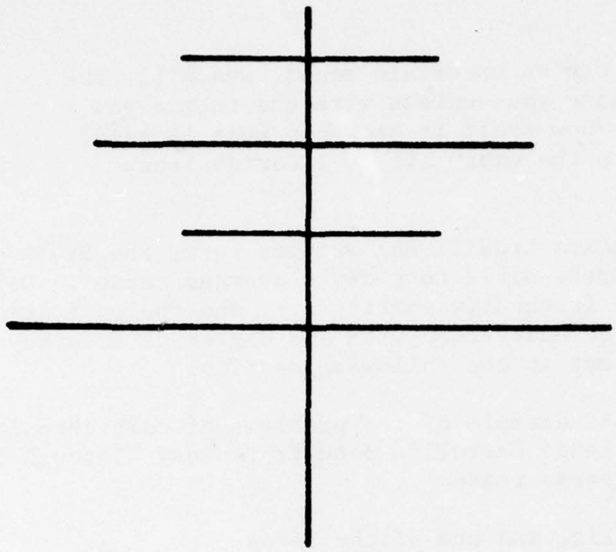

Figure 1a

Figure 1b

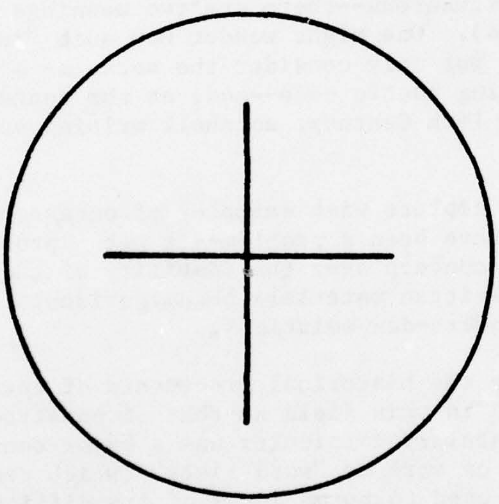

Figure 1c

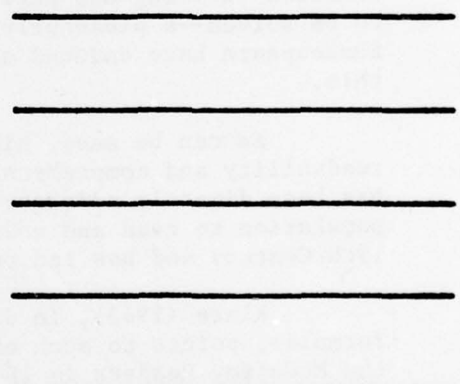

Figure 1d

Figure 1. Symbols of importance in the itinerant population.

If the trumpet give forth an uncertain sound, who will prepare for battle? So likewise you--unless with the tongue you utter intelligible speech--how shall it be known what is said? For you will be speaking to the empty air. (1 Corinthians, 14:8-9)

Yet it is quite clear that, in its traditional written form, the Bible is not easily readable or comprehensible to today's average reader. Only the nature of its content, its frequently poetic form, and the motivation of the audience permits it to be understood with any degree of accuracy. More will be said on this subject in the following section.

As another light-hearted example of the problems of comprehensibility of the written word, consider Lewis Carroll's popular fantasy "Through The Looking Glass." One familiar verse reads:

Twas Brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

Like a code, this verse is comprehensible only to the selected few who hold the key--for the most part those who have read the conversation between Alice and Humpty Dumpty: "'When I use a word, Humpty Dumpty said, in rather a scornful tone, 'it means just what I choose it to mean--neither more nor less . . . it's like a portmanteau--there are two meanings packed up into one word.'" (Carroll, 1954). One might wonder how such "incomprehensible" writing has persevered. But only consider the verse as a puzzle to be solved--a pleasantly motivating poetic code--and, as the sonnets of Shakespeare have endured since the 16th Century, so shall writing such as this.

As can be seen, history is replete with examples of occasions where readability and comprehensibility have been a problem--if not a problem that has been directly addressed. More concern over the inability of the general population to read and understand written materials became evident in the 19th Century and has led to some modern-day solutions.

Klare (1963), in discussing the historical precedents of readability formulas, points to such early work in this field as that of construction of the McGuffey Readers in 1840. Vocabulary difficulty was a major consideration in this effort. A great deal of work on "word lists" (which considered the frequency with which a word is used to be an index of its difficulty) was carried out by subsequent workers in this field, culminating with Thorndike and Lorge's work from 1921 to 1944. In 1944 Thorndike and Lorge published their "Teacher's Word Book of 30,000 Words," which has become a very useful tool in assessing writing difficulty.

Early in the 20th Century, paralleling the development of word lists, the concern became centered on the assessment of difficulty, and a multitude of "readability formulas" was constructed. These formulas and developments

beyond their relatively superficial measurement of difficulty will form the basis of much of the remainder of this report. The audience for which material is intended remains a major focal point in all this discussion. Hamilton (1970) pointed out that "Information communicated from one human to another has always needed a design or plan." We interpret the "design" or "plan" here to include a systematic means of making any writing comprehensible to the intended audience.

Contemporary Problems in Readability and Comprehensibility

There is widespread interest today in what is commonly referred to as a "mismatch" between the difficulty of much written material and the reading abilities of the users of that material. As one example of this concern, let us refer back to the discussion of the Bible undertaken in the previous section.

A great deal of concern has been expressed by clergymen and Biblical scholars to the effect that the Bible is thought of today as merely an historical document, when in fact its lessons should apply directly to modern life. This concern has spurred a number of revisions of the Bible to conform to the language usage of the present without destroying the content or the poetic form of the traditional Bible. It should be said that the traditional form of the Bible is moderately readable as measured by readability formulas. That is, providing that periods and other punctuation are made to conform to modern usage, it uses generally short words and short sentences. The comprehensibility of the traditional form is a quite different matter. Words and sentence structure familiar to the audience extant at the time of its writing are in large part totally foreign to today's reader. One source, for example, has stated:

> The Bible as a whole is essentially folk literature in which the speakers and writers are at one with their audience, expressing the interests and collective wisdom of their group--at first, the larger group of the whole Jewish nation, later the smaller group of Christian disciples. (Bates, 1951)

This statement is indicative of the nature of the intended audience of the traditional biblical form; it is clear that the characteristics of the audience have changed over the intervening hundreds of years. The same source goes on to say:

> Little can be done to revive the reading of the Bible so long as it is presented in the traditional manner. Certainly, no literary format was ever less conducive to pleasure or understanding than is the curious and complicated panoply in which the Scriptures have come down to us . . . the time has surely come when a purely literary edition of the Bible is called for: one in which the sole care shall be to present the text in as enjoyable and understandable form as possible . . . an edition for scholars is one thing; an edition for readers is quite another. The latter should make use of the former, but surely should not parade them. (Bates, 1951, p. x)

On the basis of arguments such as these, the traditional Bible is being rewritten using simplified modern language. To exemplify the results of these efforts, note this modernized version of the quote from St. Paul (seen in its original form on page 3).

> If the man who plays the bugle does not sound a clear call, who will prepare for battle? In the same way, how will anyone understand what you are talking about if your message by means of strange tongues is not clear? Your words will vanish in the air! (American Bible Society, 1971)

Another quite recent area of concern in this field is the inability of the average man to read such forms as insurance policies, income tax forms, and various kinds of contracts. The Pennsylvania State Insurance Department (as one example) recently asked companies operating in Pennsylvania to rewrite their policies to make them easier for the layman to understand (Scientific Time Sharing, Inc., Note 1). It is not known to what grade level the policies were to be oriented.

Professor Ted K. Kilty (UPI, Note 2) studied the readability level of a variety of everyday reading materials. He found, among other things, that the 1975 Federal Income Tax forms averaged about grade 9.3 in difficulty. Thus, since the estimated median reading level of the general population is approximately grade 9, perhaps as many as 50 percent would be unable to read these instructions. Another interesting finding reported by Kilty was that the reading level for the directions on a particular brand of aspirin was about 10.3--clearly higher than could be understood by many people.

In another area, New York's First National City Bank, among the nation's largest commercial banks, was among those lenders who greatly simplified the language in their loan contracts. This move prompted the New York state legislature to introduce a bill that would make the National Bank's agreement forms standard for all New York banks. (AP, Note 3)

From these examples, it can be seen that the commercial world is expressing a determination to publish more comprehensible materials for the consumption of the intended audience--the average man.

Problem

The problem that prompted this investigation is that Navy technicians reportedly are unable to read and comprehend the technical manuals which they must use on their jobs. In terms of time and cost, this problem is more serious than those which have been discussed above. Inability to use technical manuals impacts virtually all facets of Navy maintenance. Training time must often be extended to teach men to use the manuals, resulting in greater cost to the government and less productive time for the man on the job. Inability to follow maintenance directions in the manuals may result in longer time required to repair equipment, replacement of parts which are in fact operating properly, hazards to both equipment and men, and other problems, all of which affect in one way or another the Navy's operational readiness.

Objective

The objective of this report was to address the following questions:

1. What can be done to alleviate the problems of poor readability and comprehensibility of training and maintenance manuals?

2. What is within the current state of the art which will aid in accomplishing corrective action?

3. What techniques lie on the horizon which will permit correction of existing manuals or manuals yet to be written, or both?

4. What research and development questions are indicated which will replace or supplement existing or near state-of-the-art techniques for solving the Navy's problems?

After a definitive survey of the literature and of techniques in use in both the military and the civilian worlds, recommendations will be made in direct response to the Navy's needs in this area.

PREDICTION OF DIFFICULTY OF WRITING

Prediction vs. Production

This author takes a position similar to that taken by Klare (cf., 1975), which differentiates the problem of writing difficulty into two areas--"prediction" and "production." The prediction of the difficulty of writing refers to the application of some measurement device to the material to determine approximately what level of reading ability the user must possess in order to decode the written word.

As used here, "difficulty" will be taken to involve two components--readability and comprehensibility. The prediction of readability is normally a statistical function--a count, for example, of the number of syllables or letters per word, the number of "easy" words in a passage, or the number of words in a sentence. Counts such as these, taken from criterion writing samples, are then (generally) entered into a regression equation, which results in a weighting for each. A Reading Grade Level (RGL) (or an index which can be readily converted into RGL) can then be computed for test material by entering its counts into the resulting formula.[1], [2]

Assessment of the comprehensibility of written material is a more complicated task. Comprehensibility involves properties of written material which are not as amenable to counting (at least in automated form) as word length, sentence length, etc. Good style, for example, promotes comprehensibility, as does the use of simple and familiar sentence structure. Factors such as these are probably measurable only on the basis of human judgment.

One might think of "production" of readable writing as simply the opposite side of the coin from "prediction." It must be emphasized that this is not the case. There is no formula that can be used by a writer to guide him step-by-step through the composition process such that the resulting material will be either highly readable or highly comprehensible. Rather, production involves some set of rules and guidelines that (ideally) includes those features of language that have a causal relationship to the ease with which writing will be read and understood. Bormuth (1969), in commenting on readability formulas (i.e., predictive devices), points out this distinction. He states that variables that are normally used in readability formulas ". . . are, themselves, dependent upon a number of different transformations which can be performed on the language," and that they do not stand in such a causal relationship to comprehension. In short, word length, sentence length, etc. alone cannot simply be manipulated by a writer and automatically result in material that a reader can understand. This point will be amplified in the later section devoted to the production process.

---

[1]RGL is comparable to, but not equivalent to, school grade; it is a somewhat arbitrary level at which a particular grade student should be able to read with satisfactory comprehension.

[2]A detailed discussion of selected readability formulas will be given in a later section.

Terminology

The above statements--and indeed, the overall body of literature that forms the nucleus of this work--point out clearly the need for clarification of the terminology used in this subject matter area. The literature is rife with contradictions. Each author seems to assign his own unique definition to the terms he uses. This greatly complicates the already difficult task of bringing together research findings and theoretical discussions from an extensive body of writing into a coherent whole. Five concepts are particularly at issue: readability, readability formula score, comprehensibility, comprehension, and usability. For the purposes of this report, distinctions (albeit somewhat arbitrary and perhaps fine-line at points) will be made among all of these terms. This section provides the basic definition for these concepts, while the following section discusses some of their critical interrelations.

Klare (1963), a pioneer in the field, has defined "readability" as being used to indicate legibility of either handwriting or typography, ease of reading due to either the interest-value or to the pleasantness of writing, and ease of understanding or comprehension due to the style of writing.[3]

As a formal definition, and one that would be generally accepted, this would seem to be sufficiently comprehensive. It comes extremely close, however, to what is currently accepted as the definition of "comprehensibility." Another definition, offered by Sticht (1975), leaves the distinction between these two concepts even less clear. He states, "The term 'readability' refers to the comprehensibility of a publication--that is, how easy it is to read and understand it" (p. 50). Two critical questions arise. First, if, in fact, we mean virtually the same thing when we speak of "readability" and "comprehensibility," why not clear the air by collapsing the two into a single concept? On the other hand, there appears to be an operational distinction between the two. A meaningful clarification of this distinction may be established if we take "readability" to be equivalent to a "readability formula score." By doing so, the operational definition of readability is accepted as "what a readability formula measures." Such a definition excludes certain subtle aspects of good style, the smooth flow of ideas, the use of familiar sentence structure, etc., because formulas simply cannot assess these factors. They are characteristics of comprehensibility, and as such, probably can be assessed only by a human judge. Let us illustrate this distinction further. Consider the nursery rhyme "Mary Had A Little Lamb." Using the operational definition of readability suggested above, the readability of the rhyme is very high (i.e., easy). In this instance, the comprehensibility of the writing is also very high, due to the simplicity of style, its smooth flow of ideas, etc. But consider what occurs if the sentences in the rhyme are left intact but the words within each sentence are "scrambled" in some fashion. The readability, according to formula, is identical to the original version; the comprehensibility, on the other hand, is no doubt considerably lowered. Suffice to say that, in writing in which the author uses care to write smoothly and avoids unnecessarily long words and sentences, the readability score and the

---

[3]Klare later (1975) distinguishes between such a "formal" definition and an "operational" one, a point to which later reference will be made.

comprehensibility of the work will certainly be correlated. One must keep in mind, however, that high readability does not ensure high comprehensibility, and that present-day formulas cannot assess comprehensibility. It would seem a prudent rule to say that high readability is a necessary, but not a sufficient condition for high quality writing.

Both readability and comprehensibility are properties of the written material. Comprehension, on the other hand, is a property of the user. No matter what the indices of readability and comprehensibility may be, a certain level of reading skill and general intellectual ability is required in order for the man to decode and act upon written instructions. It is difficult to precisely define the concept of comprehension. In 1917, Thorndike described reading comprehension as ". . . a very complex procedure, involving the weighing of each of many elements in a sentence, their organization in the proper relations to one another, the selection of certain of their connotations and the rejection of others, and the cooperation of many forces to produce the final response" (Thorndike, 1917, p. 323). We may not have made much progress since those words were written. As aptly stated by Simons (1971), "Thorndike's description still almost exhausts the accumulated knowledge of this fundamental process." We have at least, however, developed primitive models for a better understanding of the variables involved in comprehension. These will be touched upon in the following section.

The fifth concept which seems to cause some confusion in the literature is usability. Like readability and comprehensibility, this characteristic is also a property of the material, but in a somewhat different way. It involves consideration of additional factors. Given that the material is both readable and comprehensible, can a man effectively use it on the job? This depends in large part on the characteristics of the job itself: Is the work space cramped or spacious? Is it well or dimly lighted? Is it exposed to the elements? Is the nature of the task one of troubleshooting, routine maintenance, or otherwise? The answers to these questions, and others like them, determine the usability of a given technical document. That is, in addition to the use of familiar words, short sentences, smooth style, simple sentence structure, etc., usability involves the media by which the information is presented to the technician. One might well consider that as READABILITY is a necessary but not sufficient condition for high COMPREHENSIBILITY, both readability and comprehensibility are necessary but not sufficient conditions for high USABILITY. A detailed discussion of the area of media is beyond the scope of this paper. Suffice to say that, regardless of the media employed to transmit information, it will not be "usable" if the conditions of readability and comprehensibility are not first met.

Readability, Comprehensibility, and Comprehension

Because of the integral nature of these three concepts with regard to the present work, an examination in greater depth seems warranted. To repeat, the mere fact that a passage or sentence is readable (according to formula score) and comprehensible (according to human judgment) does not ensure its comprehension. This can be illustrated directly on the basis of the passage below.

The procedure is actually quite simple. First you arrange things into different groups. Of course one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities, that is the next step. Otherwise you are pretty well set. It is important not to overdo things. That is, it is better to do too few things at once than too many. In the short run this may not seem important but complications can easily arise. A mistake can be expensive as well. At first the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity for this task in the immediate future, but then one never can tell. After the procedure is completed one arranges the materials into different groups again. Then they can be put into their appropriate places. Eventually they will be used once more and the whole cycle will then be repeated. However, that is part of life. (From Bransford & Johnson, 1973, p. 400.)

The readability of this passage (as measured by a formula called the "Fog Count") is at approximately the 6.5 Reading Grade Level (RGL). Its comprehensibility is judged to be moderately high--thoughts are in sequence, the sentences are simply constructed, the style is generally "smooth," etc. But, do you, the reader, understand what the writer intends to communicate? For many persons, there is no true comprehension of this passage, despite the fact that its level of readability and comprehensibility are quite satisfactory. In this case, what seems to be the major reason for lack of comprehension is the absence of a context, schema, or organizer for the passage. Add a context (e.g., WASHING CLOTHES) and comprehension for one of average reading ability or above becomes a simple matter. It seems clear that, in order to comprehend material, one must have some idea as to what he is reading <u>about</u>. The same idea can be applied to individual sentences. Does the reader "comprehend" the sentence: "The shooting of the hunters was terrible!"? Have "hunters been killed," or are they simply "bad shots"? These examples illustrate the deleterious effect of lack of context upon comprehension. They point to the necessity to provide, in technical materials, appropriate foreknowledge of exactly what the writing is intended to convey--spatially or temporally--with regard to the equipment in question. Readability formulas do not take the problem of concept clarity (i.e., ambiguous sentences, ambiguous topics, or poor ordering of sentences) into account. And objective judgments of comprehensibility, examining sentences or even passages in isolation, cannot directly and solely predict comprehension. A formula can index sentence difficulty under normal circumstances, but that index will be the same no matter where that sentence appears in the material. This is a problem in any sort of writing, but is no doubt more critical in technical documentation, where speed of comprehension (to say nothing of sequence) is more important than in textbooks (for example). Bransford and Johnson (1973) point out the general nature of this problem. They indicate that ". . . the differential difficulty of easy and hard items [is] not simply a function of the sentences <u>per se</u>, but rather [is] a function of the ease with which subjects [can] find solutions for the comprehension problems they represent."

To summarize the importance of maintaining clear distinctions among the various terms used in this field, one should probably consider them as lying along a dimension. Readability assessment provides a rough estimate of whether or not material will be comprehended; it is a necessary, but not a sufficient condition for the ultimate goal of comprehension. Those factors which cause differences in comprehension are not necessarily the same factors that we refer to as "readability variables" (e.g., word familiarity or length, sentence length, etc.). If material is "readable" (according to a readability formula), there is a high probability, given that the author has attempted to arrange ideas in a logical fashion, that it will also be "comprehensible." But the entire concept of comprehensibility must be qualified: the material is comprehensible to <u>whom</u>? In other words, the audience must be taken into account. Comprehensibility also is a necessary, but not a sufficient condition for comprehension; given that material is comprehensible (i.e., geared to the background knowledge and motivation of a specific audience), it probably will be comprehended by that audience.

Examination and Comparison of Selected Readability Formulas

The work with readability formulas dates back to about 1920. Klare (1963) reviewed the field of readability measurement comprehensively from the earliest work up to 1960. At that time, he reported that 21 readability formulas and 10 variations of these formulas had been published. The same author later presented an overview of formulas developed after 1960, along with modifications of certain pre-1960 formulas (Klare, 1974-1975). The majority of these formulas used as variables the length, familiarity, or concreteness of words; the length, complexity, or abstractness of sentences; or other such structural variables in a variety of combinations. Of the formulas developed to date, only six will be discussed at any length in this report. These are:

1. The Flesch Reading Ease (RE) Formula.

2. The Dale-Chall Formula.

3. The Bormuth Formulas.

4. The Harris-Jacobson Formula.

5. The U.S. Air Force "Fog Count."

6. The U.S. Army "FORCAST" Formula.

Of all the formulas developed to date, the Flesch RE formula and the Dale-Chall formula are considered to be the best validated and consistently most accurate in the prediction of readability, and thus the most widely used. The two tend to correlate very highly (normally above .90) with one another in prediction, and yet they differ in at least one major characteristic. The RE formula uses both sentence length and word length (number of syllables per 100 words), with weightings determined (as in most formulas) by a multiple regression equation. The Dale-Chall, on the other hand, uses only one of these structural variables directly--average sentence length. As its other variable,

it uses the percentage of words in a passage that do not appear on a list of words termed the "Dale List of 3000 Familiar Words." This list (consisting actually of 2946 basic words) contains words which 80 percent of fourth graders questioned indicated that they "knew" (Dale & Chall, 1948). The Dale-Chall Formula, therefore, might be considered to use word familiarity as a variable rather than word length per se, although the "easy" words in our language do, of course, tend to be relatively short. This relates closely to the problem of technical terminology in Navy writing. In technical writing, the length (or number of syllables) in a word may have little relation to the comprehension of that word by a technician. Thus, a formula which uses word length directly as a variable may "inflate" the RGL of technical materials. The Dale-Chall Formula, with its "look-up" list of easy words, may provide a model for the solution of this problem. The present list could be supplemented with "hard words" (defined as words of three or more syllables), which are generally familiar to the intended audience, and which, if found in the materials, would be weighted less heavily in determining the RGL. A number of computer programs for the Dale-Chall Formula are available and, with minor modification, could be employed for this purpose. This problem will be discussed in more detail and in a different context below.

Another major problem (related to that of word length) with the majority of readability formulas is that they have been validated against groups of school children, or at best, civilian adults. In an attempt to alleviate this problem, Kincaid, Fishburne, Rogers, and Chissom (1975) recalculated the RE Formula using Navy personnel reading Navy training materials. There is good reason to believe that a military man whose reading test scores indicate that he reads at, say, the fifth grade level, is far different in his abilities from a 10-year-old child who also reads at the fifth grade level. While this contention has not been thoroughly researched, the fact that maturation and the accumulating needs of adulthood, to say nothing of exposure to military life, require an individual to use and recognize new words is at least intuitively obvious.

Another factor that must be considered is that reading, like other skills, may deteriorate from "disuse." The older person's reading level may have declined since his school years if that person simply did not read in the interim. In the case of an electronic technician, for example, it is very possible that such words as "transmitter" and "receiver" (both of which inflate the "difficulty" of material when word length is a variable) are more readable than say, "triumph" or "receipt," which do not inflate the RGL. The recalculation of the RE Formula with military materials and subjects should result in this formula's being more useful for TM purposes.

Each of the remaining formulas is unique in some fashion. The Bormuth Formula and the Harris-Jacobson Formula both examine psycholinguistic variables (in addition to the common structural variables) in their derivation.[4] Both consider, either directly or indirectly, the problem of familiar long words and their influence on difficulty. Both make use of a computer-stored dictionary of words. Harris and Jacobson, in addition to having compiled what is no doubt

[4]Both Bormuth and Harris and Jacobson developed multiple formulas for different uses, but only one from each source will be considered in this report.

the largest dictionary of "common" terms, also constructed a list of "technical" terms for the subject matters of physics and chemistry. The Fog Count and FORCAST Formulas are unique in that both were developed using military personnel. The Fog Count was developed originally for the U.S. Air Force by McElroy (1953) and was later renormed on Navy enlisted personnel using Navy rate-training materials, at least some of which were technical in nature (Kincaid, et al., 1975). The FORCAST Formula was developed using Army personnel and Army literature (Caylor, Sticht, Fox, & Ford, 1973). These authors, in discussing the limitations of existing formulas, pointed out that a readability index was needed ". . . that would be (a) based on essential Army job reading material, (b) normed for the young adult male Army recruit population, and (c) simple and readily applicable by standard clerical personnel without special training or equipment." Each of these four formulas will be discussed in detail below.

Bormuth (1969) undertook a very extensive correlational analysis of several hundred variables dealing with sentence syntax, vocabulary, inter-sentence syntax, and linguistic variables in writing. His interest was two-fold, with both objectives being oriented toward increasing the effectiveness with which students acquire knowledge from written instructional materials. First, he was concerned with analyzing the processes involved in language comprehension. Second, he was interested in identifying those linguistic features that cause the language in instructional materials to vary in comprehension difficulty. A large number of the variables considered showed moderate to high correlations with his criterion variable (scores on a "cloze" test) and also high correlations with each other. The Bormuth formula that has been selected for discussion here is intended for computer application. It uses as comprehension variables letters per word, words per sentence, and the average number of words per passage which appear on the Dale List of 3000 Familiar Words. These variables were selected by virtue of the feasibility of obtaining the necessary data by computer scanning of the material. Bormuth indicates that certain psycholinguistic variables, such as the use of relative clauses and prenominal adjectives, could well be very important to comprehension, but are not feasible for automation at this time. The results of his work offer a number of stimulating ideas for future research dealing with psycholinguistic variables affecting difficulty. Bormuth's "machine computation formula" for determining passage difficulty is shown below:

$$\begin{aligned} \text{CLOZE MEAN SCORE} = {} & .886593 - \\ & .083640 \text{ (LTR / WD)} + \\ & .161911 \text{ (DALE LONG-LIST WORDS)}^3 - \\ & .021401 \text{ (WDS / SENT)} + \\ & .000577 \text{ (WDS / SENT)}^2 - \\ & .000005 \text{ (WDS / SENT)}^3 \end{aligned}$$

Harris and Jacobson (1972) constructed a computer-stored dictionary of 16,849 separate words (the "H-J List") by analyzing the total vocabularies of 14 series of textbooks for grades one through six. In total, 127 books containing about 4,500,000 running words were examined. In addition to the

basic H-J List, a technical list was established for each of four subject matters (i.e., physics, chemistry, biology, and social studies). Words on these lists were identified as words not on the "core" list (which would be considered common) but which were included in each of the two series of texts in each specific subject matter. That is, the "technical" words were those that are not common in basic vocabulary, but which played a part in understanding the subject matter. The purpose of Harris and Jacobson in identifying these words was instructional. That is, each such word was presumed to be novel to one naive in the subject matter. Therefore, when it was first introduced in a textbook, it was clearly defined and in addition was used three to five times in close proximity to its definition. The assumption underlying this technique was simply that the repetition of the novel word in close conjunction with its definition would reinforce that definition, and the word could be considered learned. The problem in technical materials is different. It is not that a learning effect is expected; rather, it is that one would like to assume that the reader already knows the meaning of the word and is able to read it, and that it therefore does not contribute to the difficulty (according to formula) of the passage. It would seem to be a relatively simple step forward to use a dictionary of technical terms to this latter advantage. The Dale-Chall list or Harris-Jacobson list could be adapted for this purpose and renormed on military personnel and materials.

Harris and Jacobson, in working with school textbooks, have also extracted a number of "spelling variables," which have been found to correlate with reading difficulty. For example, they have obtained squared multiple-correlation coefficients of .78 to .92 between formulas using a vocabulary list, sentence length, and "about 10 spelling patterns" with their criterion of difficulty (Harris & Jacobson, 1975). These same authors (Harris & Jacobson, 1974) present five formulas based on five different sets of variables, each of which is intended for a somewhat different purpose. Their formula number five, containing the variables "percent of uncommon words," average sentence length, and percent of "long words," could be a possible model for a dictionary-type formula for indexing technical writing. This formula is given below:

$$\begin{aligned} \text{READABILITY LEVEL} = {} & .118\ (\text{PERCENT UNCOMMON WORDS}) + \\ & .134\ (\text{AVE. SENTENCE LENGTH}) + \\ & .032\ (\text{PERCENT LONG WORDS}) + \\ & .424 \end{aligned}$$

Turning to the Fog Count, we see that its major advantage in indexing readability is its relative simplicity for manual computation. Given a double-spaced draft of the material, the person making the count writes above each word a "1" if the word is easy (defined as one or two syllables) and a "3" if the word is hard (three or more syllables). The number of 1s and 3s in the passage is summed and divided by the number of sentences to obtain the Fog Count. The count is then easily converted into RGL. According to the Kincaid, et al. (1975) renorming of the formula with Navy personnel, one simply subtracts three from the average Fog Count and divides by two to obtain

RGL. The U. S. Air Force manual (1969) recommends that writers conform to an average Fog Count per sentence of 20 or lower. (The higher the count, the more difficult the material.) The Fog Count, although it does not directly employ a "dictionary" of technical terms, does consider a problem similar to the technical term problem, and could be adpated to help overcome that deficiency. Certain long but familiar words are given a count of 1, despite having three or more syllables, according to a fairly simple set of rules. For example, given "Sacramento, California," each word is assigned a count of 1; or, given "General Robert E. Lee," the entire title and name is given a total count of only 1 (U. S. Air Force, 1969). A similar set of rules for technical terminology could be devised such that technical terms from a preestablished dictionary are weighted lightly. However, such a dictionary would be unwieldy to use except in automated form. Although Klare, Rowe, St. John, and Stolurow (1969) included the Fog Count (and the RE Formula as well) in their Computer Aided Revising, Editing, and Translating program (CARET I), it was in its original form and not in the stored dictionary sense discussed above. Adaptation to a dictionary format would seem to be relatively simple. CARET I represents a step forward in readability analysis, even though it is not primarily intended for technical manual work. It yields a triple-spaced computer printout with the number of syllables in each word printed under that word and the number of words in each sentence printed under the first word of that sentence. Further, at the end of each passage, the RGL (based on the RE Formula, the Fog Count, and three others) is printed out for that passage. It should be noted that Klare et al. automated the Fog Count in its original form and that Kincaid et al. renormed the count with Navy personnel but did not automate it. A research direction is indicated to determine if the renormed Fog Count in automated form with a dictionary format is feasible and effective and, if so, what words would be included in the list. The following section will go into more detail regarding the very complex problem of identifying terms to be included in such a dictionary.

In the development of the final formula--FORCAST--Caylor et al. first used a modification of the RE Formula and then examined the loadings of 15 structural variables (e.g., word length, sentence length, number of one-syllable words) in a multiple regression equation. Army recruits then read selected passages from seven Military Occupational Specialties (MOSs) and were tested for comprehension. The RGLs of the subjects were determined on the basis of the United States Armed Forces Institute Achievement Test (comprehension and vocabulary). The difficulty of each reading passage was then determined by a comparison of test scores and RGL (to be used as a criterion measure). The formula was then applied to determine the weighting of each of the 15 variables derived from the passages. The number of one-syllable words correlated very highly with test scores; because it is also the easiest measure to apply, it was selected as the single variable to be included in the FORCAST formula. The simplified formula then became:

$$\text{FORCAST Readability in RGL} = 20 - \frac{\text{number of one-syllable words}}{10}$$

As shown, one simply counts the number of one-syllable words in a 150-word passage, divides that number by 10, and subtracts the resulting value from 20 (Caylor et al., 1973). The FORCAST Formula is valued for its simplicity of computation; no automation is required (although automation would doubtless be useful). It does not, however, consider the problem of long, technical terms. Since the number of one-syllable words is inversely proportional to words of two or more syllables when passge length is restricted to 150 words, the readability level of technical writing might be inflated to some degree. In addition, formulas which do not include a variable dealing with length and/or complexity of sentences do not tend to hold up well in validation.

The automation of readability assessment is of paramount importance. Technical materials are being produced at a rate that simply precludes assessment entirely by hand. For example, the documentation on the F-9F aircraft in 1950 amounted to about 2000 pages, compared to 260,000 pages of technical material associated with the F-14 aircraft in 1975. Several breakthroughs in the area of automation indicate some directions for future progress. The CARET I program discussed above offers possibilities of assessment of readability early in the TM production process. A somewhat more direct measuring device is represented by the Automated Readability Index (ARI) (Smith & Senter, 1967). Here, a tabulator is connected to a standard office typewriter (IBM Selectric), which automatically counts sentences, words, and strokes as the material is being typed. This information is then inserted into a regression equation similar to the Flesch RE Formula to determine RGL. Kincaid (1972) reported a dramatic increase in reliability of the ARI over the Fog Count, as well as a much higher degree of user preference. Kincaid also reported that the Lockheed Georgia Company found it a very simple matter to calculate the ARI (presumably meaning the RGL as well) by computer as the draft versions of "technical orders" were printed out. However, typing errors tend to have some effect on the ARI computation.

The McDonnell Douglas Corporation (Whalen, 1974) used a similar device, based on the ARI principle and Smith and Senter's original formula. They call this device the Readability Ease Assessment Device (READ). As with the ARI, READ tabulates sentence, word, and stroke count as material is being typed, but further includes a calculator that presents the RGL to the typist at any point during preparation. While both the ARI and READ devices offer efficient indication of readability during the production process, it is considered that neither inherently possesses the potential for the more sophisticated operations that are available with computerized systems such as CARET I.

Before concluding the discussion of "prediction devices," a word should be said of the use of measures other than mathematical formulas. Carver (1974) hypothesized that "teachers, or other qualified experts, might be employed to produce objective ratings of passage difficulties, which, in turn, might be more valid than the Dale-Chall or Flesch Formulas" (pp. 20-21).

He developed a method termed the "Rauding Scale of Prose Difficulty"[5] with "an anchor set of six passages ostensibly representing grades 2, 5, 8, 11, 14, and 17." Qualified experts read new passages to be scaled and determine their position relative to these anchor points. Carver found in one study (involving only six raters) that the Rauding Scale correlated moderately (.71 to .82) with the Flesch, Dale-Chall, and ARI measures on the experimental passages. It should be noted, however, that only 10 of 60 students enrolled in graduate reading classes "qualified" to accurately use the Rauding Scale (p. 22). Considering the time involved in the combination of selection of raters and their actual reading and judging of target passages, this method would seem to be questionable at best for large volumes of material. However, as Carver points out (and as the earlier discussion of judgment of comprehensibility also indicated), if a passage contains "choppy sentences and inappropriately inserted little words," an expert rating might be more valid than any formula (p. 26). This author considers the rauding procedure untenable for use other than in a research situation at the present time in view of cost-effectiveness considerations. The sheer volume of materials in the Navy would doubtless require a prohibitive number of man-hours if the rauding method were used. But it must be emphasized that human judgment may be the <u>only</u> way to ensure comprehensibility, which points to a need for further investigation of procedures such as rauding.

It can be seen that a great deal of effort has gone into the development of measures that will accurately predict the difficulty of written material. The application of existing methods to technical writing remains a matter for investigation, but advances in automation would seem to offer promise in such endeavors.

<u>Basic Issues in the Prediction of Difficulty of TMs</u>

A number of quasi-theoretical problems have been touched upon in the foregoing sections. A further analysis of these problems in the context of a research and development program proposal should put our efforts into better perspective.

It has been pointed out that a prediction of low difficulty according to formula does not provide assurance that TMs will be comprehended by their intended users. In view of this fact, one might well ask whether prediction by formula is at all worthwhile. It is maintained here that since readability, under normal circumstances, is a <u>necessary</u> (even <u>if</u> not sufficient) condition for high comprehensibility and ultimate comprehension, prediction is a far from trivial issue. A more important question, it would seem, is <u>when</u> prediction of difficulty is cost-effective and when it is not. In terms of time and cost, it would be virtually impossible to determine the RGL requirements of every technical or training manual in use today. There are, nonetheless, many materials that are being used, and will continue to be used, by personnel whose reading abilities simply do not permit them to understand readily what they are to do. Readability and comprehensibility are major factors contributing to this situation. One report (Johnson, Relova, & Stafford, 1972), studying

---

[5] The term "rauding" refers to "reading and understanding" to distinguish it from reading without comprehension.

the relation between the readability "gap" and noncompliance with Air Defense Command procedures, found that a 40 percent reduction in errors could be achieved if the gap were eliminated. (This is probably a conservative figure, at that, because of the crude data used in the analysis.) Another study (Kulp, 1974) examined performance errors, performance time, and training time for employees in a private corporation who were using documentation that exceeded their reading abilities. All three of these factors were found to increase significantly when the difficulty level of the material exceeded the reading ability of the person by 2 years or more. Using actual company estimates, the cost to correct the errors made by employees with such a difficulty-ability gap in this study alone would have been over $43,000. This cost, furthermore, does not take into account the additional training time and additional performance time required when the 2 or more year gap existed. Application of readability analyses to these materials and subsequent revision to bring them in line with employee abilities would quite likely have resulted in significant savings.

In the face of this kind of evidence, it seems obvious that the Navy must have the capability for analysis and updating of at least a portion of the manuals now in use. Furthermore, there must exist a rationale for deciding what materials shall be analyzed. A basis for such a rationale might exist in the suggestion of Duffy, Carter, Fletcher, and Aiken (1975) that a dual reading standard be adopted for Navy personnel. That is, in certain craft areas, the minimum acceptable level would be equivalent to a 9th grade RGL; in others (less critical in terms of injury or damage potential), the standard would be about 6th grade. Materials used by men in the latter category would be obvious candidates for analysis to determine if they meet the minimum RGL requirements and for subsequent revision if they do not.

A second issue dealing with the effectiveness of readability formulas for predicting the difficulty of technical writing is that most such formulas have been developed and validated using school children or civilian adults reading materials appropriate to their particular needs. Only recently have formulas been developed using military job-relevant materials and validated against military criteria.

This is a two-dimensional problem. First, there is the general concern that "Reading Grade Levels" are inappropriate for identifying the abilities of military personnel. To repeat an earlier contention, it seems clear that a "fifth grade reading ability" for a child of 10 years and for a young man of 20 does not really mean the same thing. While there is little or no concrete evidence to support this contention, it is intuitively obvious that the latter, by exposure to such tasks as application forms, driving tests, income tax returns, etc., has accumulated a store of recognizable words that are unknown to the child. At the same time, the school child may very well be able to read words that the man cannot. Since reading is a skill which can easily deteriorate through lack of use, the child who is reading something each day may well recognize words that have disappeared from the older person's repertoire. This problem is deserving of attention in all types of writing for the military, including the construction of reading tests themselves.

A second, and closely related, problem has to do with the effect of technical terminology on the prediction of readability. In the current work, the concern is with technical writing for the Navy technician. The large majority of readability formulas treat the number of syllables in words as contributing to difficulty. This parameter may be questionable for use in technical writing. For example, the word "oscilloscope" may be necessary to tell a man the job he is to do. That is our concern--the man must be able to do what is required of him. There may be no other "simple" word that can be substituted. To subject a technical writer to a specification of "avoiding the use of words of three or more syllables" is comparable to telling that writer to "write about something else." A TM for a radar system which did not include such words as "transmitter," "receiver," "ionosphere," etc. would have to be considered suspect in terms of its ability to transfer information. One qualification should be made here. The percentage of technical terms has generally been shown to be about 10 to 15 percent in typical technical writing. Traditional readability formulas are probably useful in their present form in accounting for the remaining 85 to 90 percent of words. They would simply tend to overestimate the difficulty to some extent relative to 10 to 15 percent of the total number of words. Research is clearly indicated, therefore, to determine the degree of this overestimation, and to make adjustments to formulas if it proves intolerable.

Careful examination of the passage below, taken from a sonar operator manual (about 19% of its words are "technical") indicates that, if one deleted such words as "transducer," "mechanical," "equipment," etc., it would be very difficult to convey the desired meaning. This is not to say that this particular passage is satisfactorily written. In its original form, the RGL (determined by the Flesch RE formula) is 15--taken at face value, this would indicate the need for a college-level reading ability for satisfactory comprehension.

> The transmitter is keyed automatically at preset time intervals and generates a signal-frequency pulse each time it is keyed. These pulses are fed to either a topside or bottom side mounted transducer as selected by the operator. The transducer used converts the electrical pulses into sonic pulses which are transmitted through the water omnidirectionally in a horizontal plane. Sound waves will be returned to the transducer should the outgoing pulses strike reflective objects in the water about the submarine. The transducer converts returning sonic energy into electrical signals which are utilized in the equipment to produce a polar plot, centered on own ship, of targets or noise producing objects within the area. Range and bearing of echo-producing targets and bearings of noise sources may be determined from the PPI screen. Received signals are also applied to a mechanical scanning device which is positioned to any desired bearing by a servo mechanism controlled by the BEARING CONTROL handwheel on the control indicator. Signals resulting from sound waves arriving along the line of bearing to which the mechanical scanning device is trained are amplified to produce audible sounds on loudspeaker or headset. Sounds resulting from noise or noise indications on the PPI may be weak or absent immediately following transmission due to TVG action. (From NAVSHIPS 92154.2(A), Operator's Handbook for Sonar Set AN/BQS-2 (U), p. 4.0.)

The question of dealing with technical terminology would seem to have three major aspects, which should be dealt with in an interactive fashion in any R&D effort. First, there is the question of whether or not technical terms in fact are problematic in predicting the difficulty of writing. Second, there is the problem of identifying such terms. This problem must be addressed both preceeding and accompanying the first question; that is, a tentative identification of such terms would provide a corpus with which a comparison could be made to determine the degree of effect they have on the prediction of difficulty. Then, presumably, a more thorough listing of such terms would be required if, indeed, the problem is found to be significant. Finally, formulas or other predictive devices would then have to be revised to take such terminology into account.

A more detailed examination of the latter two of these steps would seem to be in order. If the prediction of difficulty of technical material must proceed in a manner different from that of traditional readability measurement techniques, the solution may be simply to treat technical terms (once identified) in such a way that they do not "inflate" the analysis. This might be done in one of several ways. First, for a given technical field (e.g., electronics), a dictionary of those terms determined by consensus to be already familiar to the average technician could be excluded or deemphasized in computation of the readability index. There are precedents that indicate promise for this approach. Dale and Chall (1948) used the percentage of words not on the Dale list as a variable in their formula. Others (cf. Brown, 1965; Stocker, 1968, 1971-1972) have measured readability of specialized writing in a similar manner. Bormuth and Harris and Jacobson, whose formulas were discussed above, have used computer-stored dictionaries in their readability calculations. Once Navy technical terms have been identified, it would not be difficult to construct a computer-stored listing of such terms to be used in a manner similar to the above examples. If no computer-stored dictionary is available, the same technique could be employed by simply applying traditional formulas, but designating which technical terms are to be considered "simple" and entering them into the formula as such. For example, if a hand count is necessary, the word "transducer" in the above passage could be counted as a word, but not as a polysyllabic word. Or, as Flesch (1950) recommends, the term might be counted the first time it appears, but not counted in subsequent appearances.

The preceeding discussion assumes throughout that technical terms of concern have been identified. This task is not so easy as one might suppose. In the field of electronics, for example, a basic dictionary of terms could be constructed by an analysis of electronic textbooks and manuals. But to whom would this dictionary apply? Will it be sufficient to have only one such dictionary? If there is only one, will it apply equally as well to the difficulty of material designed for basic electronic training (e.g., "A" school) and the senior petty officer on the job? Considering these questions, the research involved would seem to be a difficult, but not insurmountable, effort. One possible source of an initial dictionary might be a publication by American Institutes for Research in the Behavioral Sciences (1964). This work contains a frequency distribution for 1745 words found in a sample of Armed Forces training materials, including a fairly large number of technical

words. Thus, it would seem to be quite useful for at least providing the "corpus" of words to test the first of the three questions raised above. It could then form the foundation for more complete dictionaries if the technical word "problem" indeed proves to be a problem.

Finally, a word should be said of the issue of "use" tests and comprehension tests in the assessment of difficulty (readability and/or comprehensibility). A "use" test is generally a situation wherein the subject is given both the TM and the appropriate item of equipment and then is required to demonstrate to the satisfaction of a monitor that he can perform some specified maintenance function without assistance. This is similar to the idealized TM "verification" process. There is no question that such tests provide the ultimate criterion for readability, comprehensibility, and usability. The problem of cost-effectiveness remains to be answered through R&D.

Despite the fact that work has been ongoing in the area of the readability and comprehensibility of written materials for some 50 years, there are still a number of unanswered questions with regard to its value for Navy writing. The problem of technical terminology and its ramifications must be pursued. The sheer volume of materials with which we must deal argues irrefutably for greater automation of the assessment process. The direction such automation should take is still open to question. One conclusion which does stand out is that if we are to ensure that TMs are going to perform their function in the fleet, the prediction of their difficulty, within certain boundaries, is of vital concern.

The Measurement of Comprehensibility

As was seen above, the measurement of readability is a relatively simple matter and can be automated to almost any extent. It must be emphasized, however, that no readability index can guarantee high comprehensibility. While there is normally a high correlation between readability and comprehension, we must consider a means of ensuring that this relationship exists.

One test that is becoming a common tool for the measurement of comprehensibility is the "cloze test." In such a test, passages are taken intact from the material to be tested and every n^{th} (usually every 5^{th}) word is deleted and a blank of standard size substituted in its place. A subject, without having read the intact passage, is asked to fill in the blanks with the precise words that appear in the passage. An example of two versions (words 1, 6, . . . n deleted and words 2, 7, . . . n deleted) of a cloze test using the passage from the sonar operator manual previously cited on page 19 are provided in Figures 2 and 3.

__________ transmitter is keyed automatically __________ preset time intervals and __________ a signal-frequency pulse __________ time it is keyed. __________ pulses are fed to __________ a topside or bottom __________ mounted transducer as selected __________ the operator. The transducer __________ converts the electrical pulses __________ sonic pulses which are __________ through the water omnidirectionally __________ a horizontal plane. Sound __________ will be returned to __________ transducer should the outgoing __________ strike reflective objects in __________ water about the submarine. __________ transducer converts returning sonic __________ into electrical signals which __________ utilized in the equipment __________ produce a polar plot, __________ on own ship, of __________ or noise producing objects __________ the area. Range and __________ of echo-producing targets __________ bearings of noise sources __________ be determined from the __________ screen. Received signals are __________ applied to a mechanical __________ device which is positioned __________ any desired bearing by __________ servo mechanism controlled by __________ BEARING CONTROL handwheel on __________ control-indicator. Signals resulting __________ sound waves arriving along __________ line of bearing to __________ the mechanical scanning device __________ trained are amplified to __________ audible sounds on loudspeaker __________ headset. Sounds resulting from __________ or noise indications on __________ PPI may be weak __________ absent immediately after transmission __________ to TVG action.

Figure 2. Version of cloze test where words 1, 6, . . . n are deleted.

The __________ is keyed automatically at __________ time intervals and generates __________ signal-frequency pulse each __________ it is keyed. These __________ are fed to either __________ topside or bottom side __________ transducer as selected by __________ operator. The transducer used __________ the electrical pulses into __________ pulses which are transmitted __________ the water omnidirectionally in __________ horizontal plane. Sound waves __________ be returned to the __________ should the outgoing pulses __________ reflective objects in the __________ about the submarine. The __________ converts returning sonic energy __________ electrical signals which are __________ in the equipment to __________ a polar plot, centered __________ own ship, of targets __________ noise producing objects within __________ area. Range and bearing __________ echo-producing targets and __________ of noise sources may __________ determined from the PPI __________. Received signals are also __________ to a mechanical scanning __________ which is positioned to __________ desired bearing by a __________ mechanism controlled by the __________ CONTROL handwheel on the __________ indicator. Signals resulting from __________ waves arriving along the __________ of bearing to which __________ mechanical scanning device is __________ are amplified to produce __________ sounds on loudspeaker or __________. Sounds resulting from noise __________ noise indications on the __________ may be weak or __________ immediately after transmission due __________ TVG action.

Figure 3. Version of cloze test where words 2, 7, . . . n are deleted.

Before exploring these tests further, let us use their underlying concept to differentiate against the concepts of readability and comprehensibility. The sequence of words in sentences in the passage follow the familiar syntactic rules of our language. Most persons would have very little difficulty in correctly inserting the proper word in the first blank in Figure 2. If the words in the passage were "scrambled," however (consider for example, simply the interchange of the first two words), the chances of correctly filling in the blank would significantly decrease. But, even if every word in each sentence in the passage were scrambled (i.e., if sentence length were left intact) the index of readability of the passage would not change. Clearly, the comprehensibility would decrease radically.

Research has indicated that a score of about 40 to 45 percent correct completions on the cloze test (not having read the intact passage) corresponds to a score of about 75 percent on a multiple-choice test administered after reading the passage. Since the multiple-choice test presumably measures a subject's comprehension of the passage, the high correlation between scores on the two types of tests has led some writers to state that the cloze test also measures comprehension. The position is taken here that the high experimental correlation between the two sets of scores is not a valid rationale for this latter statement. It seems clear, however, that cloze does measure comprehensibility. Recall that we have defined comprehensibility as a combined product of high readability, the use of good style, the smooth flow of ideas, and the structure of sentences. Therefore, a subject could be expected to achieve a score of 40 to 45 percent merely because the structure of the text is congruous with his understanding of the syntax of language. For example, when a noun immediately follows a blank, the omitted word will probably be an article or an adjective. A subject might simply guess what the word might be, but he most likely would use the surrounding words and the general context of the passage to narrow his decision.

Despite some lingering questions as to exactly what the cloze text measures, it has the distinct advantage of objectivity. Since the material (minus the deletions) is verbatim from the material being tested, there is no problem in selecting appropriate "foils" as in a multiple-choice test. Techniques are also available for printing the test directly by computer with appropriate blanks for deleted words. The scoring of the test (normally only the precise word or the word with minor misspellings are considered correct) is quite objective if the preselected rules are followed closely.

Further research dealing with the cloze test is indicated to answer such questions as the size of the test group that is required to obtain reliable results, the effect of technical terminology, the consideration of such terminology in selecting the population to be tested, and the percentage of the total material that should be tested to provide a good overall index. At this point, it seems safe to hypothesize that administration of cloze tests using the audience for which the material was intended (e.g., Navy technicians on the job) may be the ultimate test of the comprehensibility of writing.

PRODUCTION OF READABLE, COMPREHENSIBLE WRITING

General

Thus far, we have been discussing the PREDICTION of difficulty of written materials. A related question is that of PRODUCING readable and comprehensible writing. It cannot be emphasized enough that writing material so as to make it readable and comprehensible is not merely the opposite side of the coin from prediction or assessment. Unlike the prediction process, there is no simple "formula" to tell a writer how to perform. As Gunning (1968) so clearly pointed out: "formulas are tools, not rules . . . warning systems, not formulas for writing." Writing is an ART, but it is also a SKILL. How to "do art" probably cannot be articulated. But it is clear that the skill of writing can be improved through training and guidance. This may be of even more importance in technical writing than in other types of composition, since the author with whom we are concerned may well be primarily an engineer and only secondarily a writer. We must provide him with as much help as possible in translating his technical knowledge into a form that a nonengineer can comprehend. This help and guidance is seen to take three major thrusts: style guides (to provide the basic characteristics of good writing), military specifications (to provide specific criteria that technical writing must meet), and "tools" (to permit the writer to determine whether the critical aspects of the former are being met). These three areas will be discussed independently below, but it should be borne in mind that they interact with one another in the actual production process.

Style Guides

At first glance, style guides (job performance aids for writing, in a sense) would appear to offer the best chance for improving the quality of technical manuals. However, such guides have existed in profusion in our libraries for years (and some have even been best-sellers) and yet one is hard put to point out the benefit derived from them. The readability and comprehensibility of our TMs remains unsatisfactory in many instances. One must conclude from this that such guides are either not used by technical writers or that the information they contain is inadequate to the task of "telling one how to write." Information provided by Klare (Note 4) indicates that the problem lies, in part at least, with the latter of these possibilities. Klare reports that, in reviewing 15 source books (10 specifically for technical writing) giving suggestions for clear writing, the agreement among authors was, in general, quite low. The suggestion "use short words," which one might expect all authors to agree upon, was specifically mentioned in only two of the 15 books sampled. Outright disagreement was found for such alternatives as "be concise" versus "be complete," and "keep paragraphs short" versus "vary paragraph length." In view of such uncertainty (or at the very least lack of consensus) among "experts" on writing, it is not so surprising that technical writing tends to be unacceptably difficult. One possible solution might be to incorporate those suggestions for which consensus does exist into a single manual for TM writers. Where disagreement occurs, experimental investigation of the variables involved is clearly suggested.

Progress is being made in the area of style guides. Some recently published guides orient themselves more specifically to the problems of technical writers and attempt to incorporate what little research evidence exists into their recommendations.

Klare, Rowe, St. John, and Stolurow (1969) offer some suggestions that, although general in nature, in that they are not intended solely or specifically for technical writing, do cover a wide variety of problems. For example, in this work, considerable attention is paid to the characteristics of the user. It is assumed that this information--motivation, interest, level of education, etc.--is vital to the writer in determining whether or not his product is suitable for his reader. These authors (Klare et al., 1969) present a very useful table (after Fry, 1968), dealing with the average number of syllables and average number of sentences per passage, which is included here as Table 1. An interesting facet of this table relates to the type of changes that typically occur in the transition from about 6th grade level to a higher difficulty. Intuitively, one might think that sentence length might be the critical variable in this transition. In fact, however, the evidence seems to indicate that length of word is a more salient factor. When one thinks of the curricular changes that occur shortly after the 6th grade, this becomes clear. After this point, the student begins to acquire a new vocabulary from a large variety of specialized subject matters (e.g., physical science, history, social science, etc.), and these "technical" words tend to be longer. Therefore, sentence length is likely to be of less criticality than is the word length variable. The variation in sentence difficulty above the 6th grade is, therefore, considerably less than that of word difficulty. (It should also be noted that in word length/sentence length readability formulas, word length virtually always accounts for the greater amount of variance.) From the point of view of the technical writer, Table 1 should signal a caution. Small changes (notably in word length in the lower grades and sentence length above about 6th grade) <u>appear</u> to make large differences. But these differences are deceiving. Working from this table, for example, a change of average number of syllables per word from 1.30 down to 1.26 would indicate a change of difficulty of <u>two</u> grade levels. In fact, however (particularly if only one of the two variables is changed), the revision would probably not be that much easier to understand. In other words, changing these index variables is not sufficient; changes in <u>causal</u> variables (e.g., word concreteness or meaningfulness, or sentence complexity) are needed if the reader's comprehension is to increase comparably with formula score.

Table 1

Typical Values for Average Word and Sentence Lengths for Various Grade Levels of Writing, After Fry (1968)

Grade Level of Writing	Typical Average Number of Syllables Per 100 Words	Typical Average Number of Syllables Per Word	Typical Average Number of Sentences Per 100 Words	Typical Average Number of Words Per Sentences
1st	120 or fewer	1.20 or fewer	14.3 or more	7.0 or fewer
2nd	120 - 121	1.20 - 1.21	11.0 - 14.3	7.0 - 9.0
3rd	121 - 123	1.21 - 1.23	8.7 - 11.0	9.0 - 11.5
4th	123 - 124	1.23 - 1.24	7.5 - 8.7	11.5 - 13.3
5th	124 - 126	1.24 - 1.26	7.0 - 7.5	13.3 - 14.3
6th	126 - 129	1.26 - 1.29	5.8 - 7.0	14.3 - 17.2
7th	129 - 137	1.29 - 1.37	5.0 - 5.8	17.2 - 20.0
8th	137 - 142	1.37 - 1.42	4.8 - 5.0	20.0 - 20.8
9th	142 - 149	1.42 - 1.49	4.5 - 4.8	20.8 - 22.2
10th	149 - 153	1.49 - 1.53	4.3 - 4.5	22.2 - 23.2
11th	153 - 158	1.53 - 1.58	4.25 - 4.3	23.2 - 23.5
12th	158 - 162	1.58 - 1.62	4.1 - 4.25	23.5 - 24.4
College	162 - 169 or more	1.62 - 1.69 or more	4.0 or fewer	24.5 or more

Klare (1975), in "A Manual for Readable Writing," offers writers specific guidance that is well grounded in the research findings from both psycholinguistics and readability. Involved here is the consideration that writing will improve if the writer is presented with information concerning modifications in language that have been shown experimentally to affect comprehension. There is less of such information that one would like to have, but the psycholinguistic literature is beginning to supply certain leads. Klare clearly points out that his are only suggestions, and not rules for writing. One is again reminded that writing is an art as well as a skill; suggestions can be made to improve the skill even though the artistic aspects may remain untouched. Klare presents his prescriptions for writers in terms of word changes and sentence changes, but reminds the writer that these are not the only variables involved in making writing more readable and comprehensible. He states: "You may well need to make other changes--in organization, emphasis, inclusion of examples, etc.--in addition to those described [here]" (p. 14).

Klare gives six basic suggestions for changing words (or word groups) so as to make writing more readable.

1. Attempt to use words of high (versus low) frequency and/or familiarity.

2. Prefer shorter words (particularly in relation to "content" words), rather than longer words.

3. Attempt to use words with high association value--those which call up other words quickly.

4. Use concrete (rather than abstract) words--words which easily arouse an image in one's mind--whenever possible.

5. Use the active form of verbs rather than their nominalizations.

6. Limit the use of pronouns and other anaphora (words or phrases which refer back to previous units of material).

Klare warns, with regard to technical terms, that better substitutes may not be available for them, and, in that event, they should not be changed. Instead, they should be accompanied by their definitions, unless they are known to be familiar to the intended reader. Klare, in addition to the suggestions above, provides the user of his manual with valuable "hints" as to how the six general objectives can be met.

Parallel to the six recommendations pertaining to words, Klare also offers six readability variables relating to sentences.

1. Shorter sentences and clauses contribute to more readable writing.

2. Statements tend to produce better recall than do questions.

3. Positive or affirmative sentences are likely to be "verified" more quickly and with fewer errors than are negative constructions.

4. Active statements lead to easier recall than passive statements.

5. Self-embedded constructions tend to be more difficult to comprehend than nonembedded statements.

6. Constructions that are high in word depth (i.e., require greater "commitments" or memory load on the reader's part) tend to be more difficult than low word-depth sentences.

Klare poses three general questions to a writer with regard to these six variables:

1. Is the sentence long and complex?

2. Will intended readers understand the sentence?

3. Is the variety needed, or should I simplify the sentence?

If changes are called for, those above should be considered.

In addition to Klare's 1975 Manual (and the suggestions from Klare et al., 1969), two manuals have been written expressly for Armed Forces instructional and/or technical materials. One of these (Post & Price,

1974) was written with Navy technical manuals in mind. While there are some points in this manual that tend to contradict traditional readability terms and concepts, the basic guidance offered by the manual is clear and concise. For example, they present a series of seven key "tests" which, if applied to technical writing, should achieve better readability/comprehensibility. These tests (abbreviated to an extent) are as follows:[6]

1. Heading Review: Do approximately one-half of the subparagraphs have headings? Is material within paragraphs consistent with its heading?

2. Topic Sentence Check: Is the heading clear? Does the heading cover about three or four topic sentences or key points?

3. Words per Paragraph Count: Do paragraphs average about 45 to 60 words? Are key points highlighted if the paragraph is longer?

4. Words per Sentence Count: Do the sentences average 20 words or fewer? Have compound sentences and complex sentences been avoided?

5. Syllables per Word Count: Does the material average about 1-1/2 syllables per word? Have short words been used whenever possible?[7]

6. Equipment Nomenclature Count: Is any unfamiliar nomenclature either defined in the text or called out on an accompanying pictorial?

7. Layout Review: Have you used double-column format? Is each graphic contiguous with the text in which it is discussed or referenced?

It is believed that this manual (which also contains tests for graphics/pictorials, format, etc., and specific guidelines for correcting deficiencies), can be of valuable assistance to the technical writer. However, it is vital that it be experimentally tested to verify this assumption. Reformatting an existing manual using this guide, testing the revision with Navy technical personnel, and, if necessary, revising the manual should be a research direction of high priority.

A style guide, or "guidebook" was also developed by Human Resources Research Organization (HumRRO) to assist writers in the development of Army training literature (Kern, Sticht, Welty, & Hauke, 1974). Despite the title, the authors clearly point out that the guide is intended for use in writing job performance aids (JPAs) as well as training materials. They emphasize the effects of style not only upon readability but also on readership (the inclination of personnel to read it). Among other things, the manual presents numerous examples of difficult material (both text and illustrations) actually taken from existing manuals and, on a facing page, a

---

[6]This manual assumes that the writer is aiming at the ninth grade level; this should be considered in interpreting the tests. An answer of "yes" to any question indicates adequate readability (at ninth grade) on that count.

[7]The authors suggest that since manuals deal with technical terms that cannot be eliminated, these terms should not be included in the count of syllables. This procedure should be experimentally tested.

rewritten version of that material. Many examples of good writing are provided for the user. Further, it presents information about the actual reading abilities of Army personnel and guidelines for identifying the intended user of a manual. This emphasis on the user would appear to be extremely valuable to the writer. The step-by-step procedures for checking one's writing, using the specially developed FORCAST readability formula, are clearly presented. As with the Navy manual (Post & Price, 1974), systematic research is needed to determine the effectiveness of this manual with actual Army writers. It would seem to be very effective--in essence a JPA to write JPAs--but concrete evidence should be obtained to support this assumption.

In general, there is consensus in virtually all such guides on certain quite general attributes of good writing. Those attributes which have particular bearing on technical writing are as follows:

1. Unity. A singleness of purpose and direction should characterize every unit (clause, sentence, etc.) of the writing from beginning to end. No irrelevant material, no aimlessness, no speculation should be included. The writer should omit all extraneous materials, logically sequence all relevant materials, and use headings to promote unity.

2. Simplicity. The approach to the subject matter must be direct; the word flow in sentences must be smooth; only words suitable to the occasion and the matter at hand must be used. Writers must be able to recognize the right word for the occasion--the most ordinary word having appropriate precision and emphasis. The writer must, above all, avoid pretentiousness.

3. Conciseness. Conciseness in technical writing comes from leaving out unimportant details and from using the correct method of subordination. This invloves not only leaving out irrelevant information, but also omitting relevant information that is not necessary for the task immediately at hand. The writer must use sentences that say all they are capable of saying (and only that much), because they are subordinate to some foregoing "directive" statement.

4. Exactness and Concreteness. Any deviation from exactitude in expression at any point in the documentation is likely to lead to an erroneous or ambiguous rendering of the entire procedure. The writer must have his facts correct, use the most appropriate words to convey those facts, and combine those words in sentences in such a way as to optimize the probability that they will be understood by the readers in the manner intended by the writer.

5. Completeness. The goal of a TM is to enable the technician to complete an assigned task. No information necessary to that end must be omitted. There must be no suggestion or implication of aspects "surrounding" that task. The writer must convey completely (and precisely) all he knows and thinks relevant to the reader's full understanding and adequate grasp of all elements involved in the problem. It should be noted that completeness comes from the writer's awareness of the reader's needs, experience, and intended purpose, and not from the writer/engineer's knowledge of the system.

6. Clarity. While clarity is a prime requisite for technical writing style, it might better be considered an end, and not a means. Without simplicity and unity, the major thrust of the material will most likely be obscured. Clarity is the composite of all the other virtues, yet greater than the sum of its parts. It results from a regular appearance of expected patterns in language. It is a product of the writer knowing exactly what he wants to say and what words and arrangements of those words would say it exactly.

7. Coherence. There must be no gaps for the reader to bridge. He must be able to move with minimal effort from sentence to sentence, paragraph to paragraph, and section to section. The writer must ensure that his sentences are unfolding the subject matter in logical or chronological order, using natural expression, and providing appropriate transitional devices between sentences, paragraphs, and sections.

8. Emphasis. Emphasis is critical to technical writing. It tells the user exactly how important particular parts of his documentation are--to his safety, the consequences to the equipment, etc. The writer must attend to such characteristics as the position of emphasis (at the beginning of sentences or in specially segregated area, etc.). From the point of view of the reader, the criticality of WARNING....CAUTION....DANGER, etc., demands that they stand out from standard procedural information.

9. Restraint. The writer must not say more than the facts warrant. He must not exaggerate. He must avoid superlatives (except on occasion with regard to EMPHASIS). He must remember that the primary objective of his writing is to tell the reader what he is to do, without unnecessary amplification. That a label on a dial is painted white may be incidental; that a label on a can is red might be a different matter.

This author, in applying suggestions from Klare (1975), Post et al. (1974), and Kern et al. (1974), revised the passage from the sonar operator manual previously quoted on page 21. This revision, resulting in a lowering of the RGL from about 15 to about 7, is presented below:

> The transmitter is keyed automatically at preset time intervals. It generates a signal-frequency pulse each time it is keyed. These pulses are fed to either a top-side or bottom side mounted transducer selected by the operator. The transducer converts the electrical pulses into sonic pulses which are transmitted horizontally through the water in all directions. If the outgoing pulses strike reflective objects, sonic waves are returned to the transducer. It converts this sonic energy into electrical signals. The equipment uses these signals to produce a polar plot. The plot is centered on own ship and shows targets or noise producing objects within the area. Range and bearing of echo-producing targets, and bearings of noise sources, may be determined from the PPI screen. The signals are also applied

> to a mechanical scanning device which is positioned to any desired bearing by a servo mechanism controlled by the BEARING CONTROL handwheel on the control-indicator. Signals arriving along this line of bearing are amplified to produce audible sounds on loudspeaker or headset. Sounds caused by noise or noise indications on the PPI may be weak or absent immediately after transmission due to TVG action.

No evaluation has been performed to determine if the above revision had any effect on comprehension of the material. The only "technical" word omitted was "omnidirectionally," and the content is believed to be essentially the same as the original version.

The rewriting of material to make it more readable is not easy, but the application of these rules tends to make the task less formidable. It would be interesting to compare this revision with a revision made by the (unknown) original author. It seems to be a major problem particularly in technical writing, that writers tend to write at about the level at which they (not necessarily their readers) read, and often at their own level of knowledge (and not that of their readers). The original author might have a much more difficult time revising his work than an independent "rewriter" or an editor. Klare et al. (1969) point up this problem. They say: "Rewriting, as well as writing, is still an art . . . the task of rewriting one's own or someone else's material to make it more readable is seldom quick or easy. It tends to take longer, and is less well liked, than a comparable first writing task." The first draft of a maintenance instruction for a technician is more than likely to contain both <u>more</u> data and more <u>difficult</u> data than is desirable. Style guides must take these problems into account, and offer specific assistance in rewriting as well as orignial composition.

<u>Military Specifications and Standards</u>

A second critical problem area related to the production of readable manuals has to do with the specifications provided to the writer by the Navy. Most military specifications (MilSpecs) and military standards (MilStds) at the present time offer a general statement such as: "the level of writing should be for a high school graduate having specialized training as a technician in military training courses" (from MIL-M-24100B, 1974). This sort of statement is deficient for several reasons. First of all, it is doubted that <u>anyone</u> can simply set out to write technical material according to such a guideline. No one even knows with any degree of exactness what the level of writing for a high school graduate <u>is</u>. Secondly, there is no guarantee that the user of the resulting manual in fact meets these qualifications. In interviews with fleet personnel (notably in the electronic rates), there is an indication that all personnel, from "strikers" to top supervisors, use technical manuals for their assigned equipment. Aside from general education, it is obvious that the "experience" qualification will not be met for certain of these technicians.

Other MilSpecs give somewhat more explicit directions. For example, MIL-M-63000C states: "Narrative text (those pages that consist of not less than 200 words in consecutive sentences per page) shall conform to the following readability standards: The average sentence length (ASL) shall not exceed 20 words . . . The average word length shall not exceed 1.60 syllables . . . The percent personal sentences (PPS) shall not be less than 15 percent of the total." The particular figures mentioned seem reasonable; if such standards were achieved, the readability would almost certainly be high. But, one might wonder if some writers read any further than where it says "200 words in consecutieve sentences." Certain persons responsible for producing technical manuals have been heard to say that, "by definition," their writing never <u>has</u> more than 200 words per page in consecutive sentences.

Clearly, better guidelines for technical writers are necessary. It is suggested that, as a beginning, MilSpecs themselves should be subjected to readability assessment. Those which have been examined by this author have been found to be excessively difficult. One sentence picked more or less at random from a basic MilSpec contained 47 words, of which 21 were composed of three or more syllables. Other examples abound. But aside from their difficulty, it seems clear that substantive information must be provided that will improve the overall quality of manuals. The "tools" that are suggested below can then be used to support the writer in attaining the requirements imposed by the specifications.

<u>Tools for the Technical Writer</u>

Given that suitable writing guides are available to the technical writer, and that the user requirements are spelled out in MilSpecs and MilStds, how can the conscientious writer assure himself that his product will be readable and comprehensible?

A basic requirement would seem to be that writers be provided with readability formulas with which they can effectively predict the readability of their materials. Two formulas that, in the opinion of this author, will be most useful for the writer, are:

1. The Renormed Flesch Reading Ease Formula (from Flesch, 1948, as recalculated for Navy use by Kincaid et al. (1974):

$$\text{GRADE LEVEL} = .39\,(\text{Average Sentence Length}) + 11.80\,(\text{Syllables Per Word}) - 15.59$$

2. The Dale-Chall Formula (from Dale & Chall, 1948):

$$\text{GRADE LEVEL} = .1569\,(\text{Dale Score}^{8}) + .0496\,(\text{Average Sentence Length}) + 3.6365$$

[8]The "Dale Score" is the percentage of words in a passage which are not on the "Dale List of 3000 Familiar Words."

Neither of these formulas requires automation for its computation (although computer programs for both are available). This permits flexibility on the part of the contracting agency or in-house writer and also contributes to Navy cost-effectiveness decisions. That is, the formulas may be used manually by a writer making a quick estimate of readability as he writes or rewrites, and may also be used in automated form when an editor or quality-control checker must analyze large amounts of material.

Note that both of the short formulas above consider sentence length. The basic difference between the two is that the RE formula uses syllable count as an index of word difficulty wheras the Dale-Chall (D-C) formula involves a look-up list or dictionary. But both formulas have consistently yielded higher predictive power when compared to other formulas, and both are recommended here because they complement one another. A given writer or editor might prefer a particular one or one might be used as a check upon the other. The D-C formula is somewhat preferred here for technical writing because it does rely on a dictionary. This capability would seem relatively easy to expand to include technical terminology that, although perhaps lengthy, is "easy" for the intended audience. This advantage is offset to some degree by the fact that the D-C formula is somewhat more difficult to compute manually. Research is needed to determine if the D-C predictive power could be increased by the addition of technical terminology to its dictionary.

It has been said that "writing to formula," while obviously resulting in a low readability score, will not ensure comprehensibility and ultimate comprehension of the material. The well-intentioned writer will realize this fact and will not simply use short words and short sentences while possibly sacrificing comprehensibility. Those responsible for the quality of the final product, however, must take care to ensure that it does not occur. The short readability formulas above must not be used as the final criteria for writing quality. Some writers might attempt to circumvent the short formula prediction by making minor changes that reduce the readability score, but that in fact might have a deleterious affect on comprehensibility. One solution to this problem is to use as a quality control criterion one of the more complex formulas, such as the Bormuth formula or the Harris-Jacobson formula discussed earlier. By the very nature of such formulas, it is virtually impossible to "write" to them. But knowledge by the writer that the variables entering into such formulas will form the judgment of the quality of his work should be of assistance to him as he writes.

Once the writer has an indication of the readability level of his material, he can determine (by reference to a MilSpec, for example) whether there is likely to be a mismatch with his readers' reading ability. If a mismatch is found, writing aids (both manual and automated) will be required. Suggestions for revising material, if included in style guides would be one of the sources for such aid. Other aids, in increasing order of sophistication, are discussed below.

Automation of the writing process (to the degree that this can be achieved) would seem to exist on at least three levels at the present time. At the lowest (i.e., least sophisticated) level, there are devices such as the Automated Readability Index (ARI) and the Readability Ease Assessment Device (READ). While essentially <u>prediction</u> devices, they do play a part in production. These two similar devices, which feature attachments to office electric typewriters, provide a count of strokes, words, and sentences as material is being typed. These variables are then entered into a modified RE formula and the RGL of the material can be obtained. Thus, a typist can determine, as the material is being prepared, what the difficulty level of the writing is. Revision can then be done as necessary by the writer or editor. The major limitation of these devices (other than a need for typing skill) is that they provide <u>only</u> readability formula data. One need only consider the vastly superior potential of typing onto computer compatible tape (from essentially the same machine which would be used for ARI or READ), to realize these limitations. They could, however, have some usefulness for the immediate needs of writers and in limited situations, until more sophisticated devices can be put into operation.

At an intermediate level of usefulness is a time-sharing system such as that developed by Scientific Time-Sharing, Incorporated.[9] By typing material into a table-top console connected to a large central computer, a variety of readability information can be automatically obtained. While this in itself is not a marked improvement over the devices discussed above, it would seem to have potential for the development of other capabilities leading to a fully integrated system of assessment and production aids. Like ARI and READ, it would solve immediate problems, but would also offer a range of other problem solving capabilities in what would seem to be a cost-effective manner.

At roughly the same level as the above time-sharing system would be the existing CARET I computer system discussed under the heading of prediction. This system not only provides a great deal of readability data but offers immediate aid to writers in the form of a computer printout formatted for efficient editorial revision. This printout is triple spaced and provides number of syllables for each word and number of words for each sentence. A writer or editor can see at a glance which words or sentences are likely to cause difficulty and make changes directly on the printout. Like the time-sharing system above, CARET I offers the potential for a number of further advances which should be of help to writers and/or editors.

A third, and most sophisticated level of automation to be discussed here is represented by the TRUMP (<u>T</u>echnical <u>R</u>eview and <u>U</u>pdate of <u>M</u>anuals and <u>P</u>ublications) system currently undergoing testing at the Naval Air Rework Facility, Jacksonville, Florida. TRUMP takes advantage of many of the more advanced characteristics of the present state-of-the-are in information processing. A major aspect of the system is its Optical Character Recognition (OCR) capability. This permits optical read-in of

---

[9]No bias for or against this particular system or other commercial systems in this report is intended. Other similar (and perhaps better) systems may well exist and not have come to the attention of this author.

information as opposed to traditional keypunching or magnetic tape processing. It is reported that OCR results in input time of up to 50 times that possible with the latter methods. Further, this system has the capability for automated merging of text and graphics--a process which heretofore has required human effort of considerable expertise. In addition, the system features automated composition and phototypesetting, which reduces the processing time even further. TRUMP presently does not employ a readability analysis, but because of its sophistication can readily accept programs for this purpose. While evaluation of the system is not complete, it seems to be very near the kind of "integral" system envisioned here. Reports recieved of prior testing indicate that TRUMP is thus far living up to its expectations.

What would those of us involved in the production of technical and training materials envision as the optimum integral system for this process? Such a system would have to offer a number of capabilities that, at present, do not exist in any single system. Figure 4 shows the types of interactions with which such a system must be able to deal. Note that there is a need for automation virtually from the moment that a writer first puts pencil to paper. All of the issues discussed earlier must be taken into account--technical terminology and alternate phrases or words when possible, the fastest and most efficient throughput time possible, the ability to deal with very large volumes of material, efficient update and/or correction, etc.

Beginning with the author's first draft, there is a need for readability analyses at its earliest stages. Intuition is a vital requirement here, and obviously cannot be automated. That is not to say that it cannot be assisted. The end of the feedback loop in Figure 4 can provide this assistance and we will return to it at a later point.

When an author has completed a chapter, section, or just a single passage in his work, his intuition may tell him that the material may be too difficult. Prediction of its readability may be indicated. The alternatives presented above offer increasing degrees of efficiency with which this can be accomplished. He could manually compute the readability index using one of the formulas presented. This could be accelerated by the use of the Navy Automated Counter developed by Biersner (1975). If he is a typist, or if he as a typist available, the use of the ARI or READ would give him a fairly fast and reliable prediction of readability. But, ultimately, it seems clear that we would wish to dismiss all these possibilities and think in terms of a computerized system. The reason for this is clear. With a computer, once material is typed it is not simply made available in hard copy, but can be stored and put to a variety of further uses. This argues for a system at least as sophisticated as the Scientific Time-Sharing system or CARET I and, ultimately, for a system with the potential capabilities of TRUMP. Such a system would involve the following characteristics at a minimum.

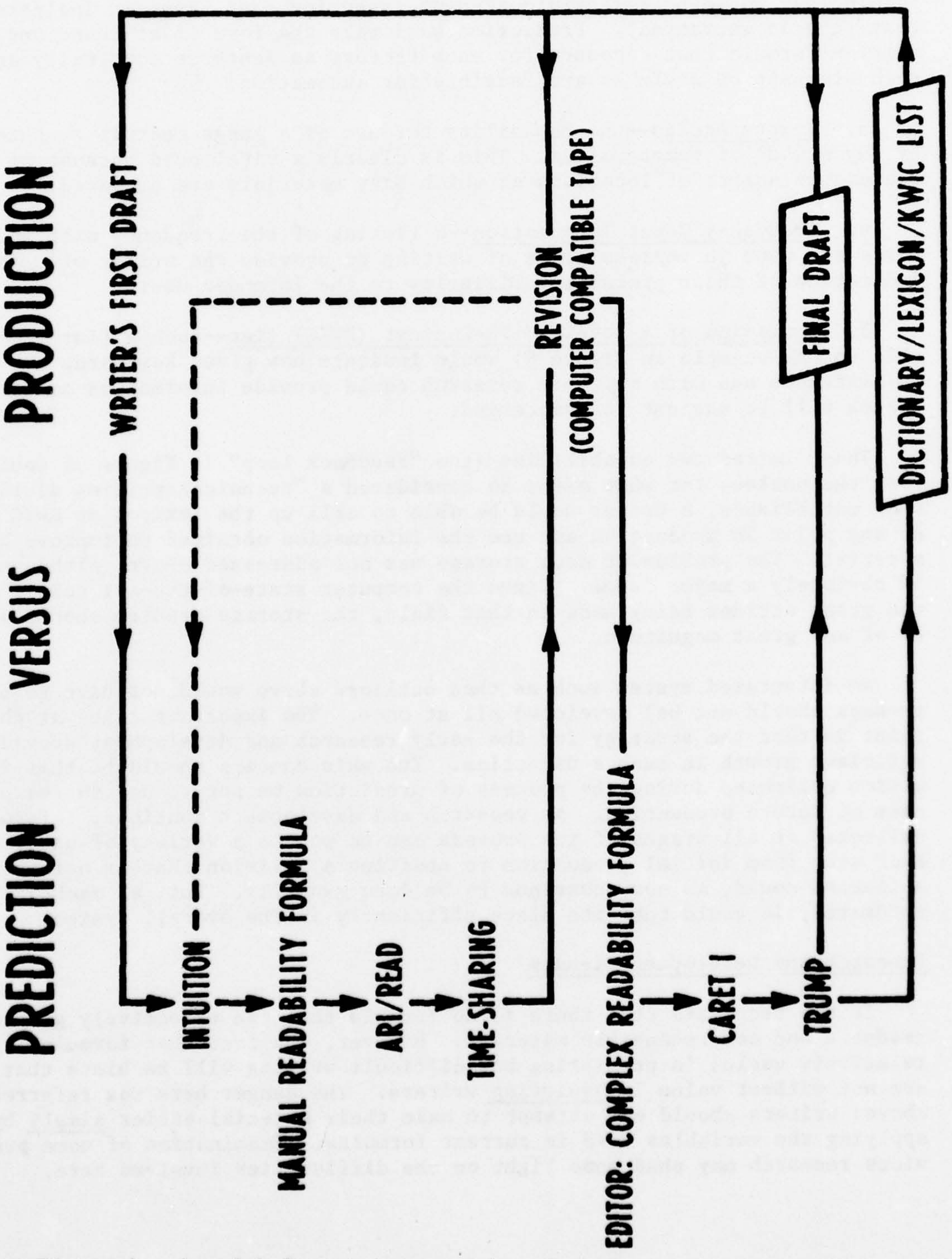

Figure 4. A model for an integrated technical manual production system.

1. Optical Character Recognition--the ability to electronically scan documents for input into computer store rather than having to laboriously key-punch the material.

2. Readability Analyses--prediction of the readability of stored information at any stage of the production process. Familiar technical terminology would be taken into account (assuming that research indicates that this is warranted). Prediction wold take the form of at least one fairly complex formula that accounts for such factors as sentence complexity and such elements of style as are feasible for automation.

3. Remote Access--the capability for use of a large central computer by any number of remote sites. This is clearly a vital need because of the tremendous number of locations at which Navy materials are prepared.

4. Frequency Count Information--a listing of the frequency with which words are used in various types of writing to provide the writer with an indication of their probable familiarity to the intended user.

5. Formation of a Key-Word-In-Context (KWIC) list--such a list (comparable to the example in Figure 5) would indicate how given key words are used in sentences and with suitable research could provide information as to which usages will be easiest to comprehend.

These latter two capabilities (the "feedback loop" in Figure 3) would form the nucleus for what might be considered a "technical writing dictionary." Once established, a writer would be able to call up the lexicon or KWIC list at any point in production and use the information obtained to improve his material. The problem of data storage was not addressed above, although it is obviously a major issue. Given the computer state-of-the-art today, and the great strides being made in that field, the storage problem should not be of any great magnitude.

An integrated system such as that outlined above would not have to be (and perhaps should not be) developed all at once. The important thing at this point is that the strategy for the early research and development provide for efficient growth in such a direction. The main concern should be that information collected during the process of prediction be put to use in the process of future production. As research and development continue, information collected at all stages of the process can be put to a variety of uses. Each step from initial production to updating a revision that is not yet automated could, as now, continue to be done manually. But, as each is automated, it could take its place efficiently in the overall system.

Research and Development Issues

It has been said that there is no formula that can effectively guarantee readable and comprehensible material. However, the fact that formulas are relatively useful in predicting how difficult writing will be hints that they are not without value in assisting writers. The danger here was referred to above: writers should not attempt to make their material easier simply by applying the variables used in current formulas. Examination of some previous research may shed some light on the difficulties involved here.

Navy technical	manual system	92154.2 4 4-3*
When under	manual operation there is danger	92146A 3 3-8
Turn control switch to	manual operation before aligning	92164.1 3 3-3
When in	manual mode PPI scope will appear	92164.1 4 4-5
Always double-check	manual computation of range data	92222A 5 5-1
The training	manual provides details for operation	92222A 5 5-2
If the figure cannot be found in the	manual index refer to operator	92286A 1 1-3
Operator	manual for further guidance	92286A 1 1-3
System provides for	manual switching of antenna	19712B 2 2-2
Look in maintenance	manual for location of power	19712B 3 3-2
	etc.	etc.

*Code indicating source, section, page, etc.

Figure 5. Example of key word in context (KWIC) list.

There have been perhaps two dozen studies that show clearly that experimental manipulation of writing variables can make a significant difference in comprehension. In general, these experimenters (many of them Ph.D. candidates) have taken original passages and written easier and/or harder versions as indexed by readability formula scores. In examining these sources (many of which are abstracted in the annotated bibliography of publications examining comprehension variables), the problem is one of specifying precisely what was changed to make the readability different. Most researchers have reported that they modified word and sentence difficulty and little else, presumably because these are the two variables common to most readability formulas. Most writers would argue, though, that such changes are either not sufficient or are too simple-minded for such a complex task. And more than a dozen other experimental studies bear them out in that they showed nonsignificant differences in comprehension of altered material that (according to formula) was "harder" or "easier." Yet the authors of these latter studies appear to have changed essentially the same variables--word and sentence difficulty. It seems clear that changes in such surface variables as word length, sentence length, etc. may incidentally result in changes at a deeper level which may or may not be apparent in formula scores.

One example may help to clarify this point. Consider the following sentence:

> The Officer of the Deck waved frantically to the helmsman and ran rapidly back to the Captain's sea cabin.

If this sentence were to occur in a passage which was being rewritten to achieve a lower readability formula score, there might be a temptation to remove the conjunction, resulting in two shorter sentences. The result could be as follows:

> The Officer of the Deck waved frantically to the helmsman. He ran rapidly back to the Captain's sea cabin.

When divided in this way, the ambiguity becomes apparent. Such ambiguity could well result in poorer comprehension despite the lower readability score. Of course, the meaning can be made clear with the addition of more words to the second sentence, such as beginning the second sentence with the word "then," or repeating the words "Officer of the Deck." There is a tendency, however, for writing to become choppy with the repetition of words. In any event, a simple mechanical shortening of words and sentences, while "fooling" most readability formulas, will <u>never ensure</u> better comprehension.

A second major issue with regard to research and development in this area also emerges from a review of the experimental literature. It is simply this. Successes and failures to produce changes in comprehension may be a function of something entirely independent of the way the material is rewritten. Klare (1975) has proposed a model that illustrates the role that other potential variables may have in comprehension. This model shows reading performance as a function of <u>reader motivation</u> acting over <u>reading time</u> interacting with <u>reader competence</u> interacting with the <u>readability</u>

level of the material. The discussion above considered only the readability level, with varying degrees of reader competence being assumed. Klare suggests that the interactions of this model make the experimental investigation of reading performance extremely difficult. He states that "the experimental situation tends to raise reader motivation (and experimenters often take specific steps to achieve this). If this extra motivation has an opportunity to take effect (through provision of generous reading time, for example), experimental differences tend to be washed out" (p. 6).

As the model suggests (and is borne out in much of the experimental literature), motivation is clearly a factor in readability and comprehensibility research. Work is definitely needed to determine systematically how this factor operates and how it can be manipulated to best advantage. Klare (Note 4) describes four categories of variables that enter into this problem in an interactive fashion. First, there is the choice of subject matter. Choosing content in which the reader already has a high level of advance knowledge or that has a high interest value for the reader will tend to decrease the likelihood that readability *per se* will have a significant experimental effect. In other words, as knowledge and interest in a subject matter increase, the importance of readability decreases. Choosing content for which the converse is true (low background information and/or low interest value) will tend to increase the probability of obtaining significant differences between more readable and less readable versions of material, and should, therefore, be given first priority in any revision process. A second variable has to do with the treatment of the independent variable, i.e., readability. For example, the research literature indicates that changing only one readability variable (e.g., words or sentences) will decrease the probability that readability will make a difference in comprehension, while changing both of those factors will increase that probability. Another aspect of this issue is the amount of difference (gap or mismatch) between difficulty level of material and the RGL of subjects. A small experimental difference or mismatch will tend to decrease the likelihood of significant findings, while a large difference will increase it. A third variable in readability experimentation has to do with treatment of subjects. This is exemplified in the description of Klare's model above. Increasing motivation and providing an opportunity for it to take effect (e.g., by encouraging rereading) will tend to decrease the chances of obtaining a significant result which can be attributed clearly to readability changes. This combination of circumstances often occurs in experimental settings, but seldom occurs in a field (i.e., a "typical") reading situation. This appears to be a major reason why field studies show significant effects of the readability variable so much more consistently than do experimental studies. And finally, the fourth class of variables deals with the dependent variable--the test situation or the test itself by which comprehension is measured. Some of the examples of how this is seen to confound studies are: (1) use of small numbers of or unrepresentative multiple-choice items, (2) use of only one version of a cloze test (e.g., with only the 5th, 10th, . . . 25th words deleted), or (3) having the readability of the test more difficult than that of the experimental text being covered. Special attention must be given to all four of these variables if readability *per se* is to have a chance of making a difference in an experimental situation.

It should be pointed out that the foregoing are research issues to be considered in investigating readability and comprehension. Many of the recommendations made do not apply if one is attempting simply to improve comprehension. For example, if one's goal is to obtain the highest possible degree of comprehension, it would be prudent to increase motivation and allow generous reading time, rather than the opposite.

GRAPHICS AND PICTORIALS

General

The area of comprehensibility of graphics and pictorials is a much less well-developed one than the area of text readability and comprehensibility. Means of coming to grips with the problems of the area are elusive. The major reason for this parallel the area of text: the parameters that contribute to the comprehensibility of tables, charts, schematics, etc. are for the most part unknown. It is difficult even to identify candidates for such parameters in any number. Other problems that contribute to the difficulty of examining the area of graphics include: (1) the large number of different forms, or "tokens," under the rubric "graphics," (2) the inconsistency of terminology for graphics in the literature, and (3) the difficulty of determining the comprehensibility or usability of any given graphic in isolation (i.e., without accompanying text).

Ideally, one would like to analyze graphic material and attribute to it some index of its difficulty for a given audience (i.e., prediction). At the same time, procedures which would simplify and increase the speed, cost-effectiveness, and quality of processing graphics are needed (i.e., production). It can be seen from this that the ideal approach to improvement of graphics is analogous to that of text. But there are a vast number of variables which are known to correlate rather highly with comprehension of text (although the combination which provides the best prediction is not established). There are only a few parameters (and these understood only in gross terms) which contribute to the comprehensibility of graphics. Density (information per unit of area), for example, certainly affects graphic comprehensibility, but optimum density is simply not known. It is clear that, if significant progress is to be made in improving technical graphics and pictorials, the first step must be to identify the basic parameters affecting their usability.

Identifying Critical Graphic Parameters

For the purposes of this report, "graphics" will be used as a generic term to cover any type of illustration found in military training or maintenance manuals. For greater clarity, however, a distinction will sometimes be made between the two major subdivisions of illustrations. The first of these are line illustrations, which consist of black and white graphs, diagrams, and schematics, among others. The second are half-tone illustrations, such as photographs and airbrush drawings. As one might expect, the preparation and printing of the latter are more expensive, more time-consuming, and more complicated than of the former. Thus, the basis for making the distinction: it may be that our immediate goals should center on line illustrations, with concern for improvement of half-tone illustrations to come later. Whenever it is appropriate in the discussion to follow, the two will be distinguished; for general purposes, however, the generic term "graphics" will be used.

In general, the literature indicates that the field of graphics in general has made huge strides over the past decade. The majority of these advances, however, have not seemed to deal with as complex a problem as the one the Navy faces. Huggins and Entwisle (1974) summarize very well what many others are also saying:

> . . . computer technology has been dominated by persons who seem to be happy with a simple, very limited alphabet of characters used to produce linear strings of symbols, and only in the past several years are graphic terminals becoming commonplace. Also, the interests and objectives of those who design and build this elaborate equipment are quite different from, say, the educator who is primarily concerned with communicating concepts . . . the need to explore this potential is great and offers many challenging topics for research." (pp. xi-xii, italics added)

In effect, what these authors and others are saying is that, despite great advances in computer technology, the concern of most researchers has been limited to relatively simple and discrete classes of data. A great deal of effort has been expended, for example, on the development of techniques for recognizing letters of the alphabet, in any variety of type fonts or even hand printing. By analogy, we require a capability of distinguishing, for example, the symbol for a transistor, and beyond that the relation of a given transistor to other elements in a schematic.

As with the area of text, it seems both desirable and economical to have a model or schema from which to examine the problems of graphics. The terms "schema" and "model" are used in the sense of explaining what a complex phenomenon consists of and of generating hypotheses for extending beyond the facts on which the model is based. One possibility for such a model is presented in highly preliminary form in Figure 6. It is immediately evident from this model that there are many branches of the tree still to be identified. This involves a search for those parameters of various kinds of graphics which promote their usability. A satisfactory set of such parameters has yet to be discovered. One possibility for this deficiency may be that a more complex model is required. The literature is replete with references to this problem. Rosenfeld (1969) for example, states:

> Given a set of properties [parameters], it is in principle possible to measure the contribution which each of them makes to the classification decision . . . . Of greater relevance to pictorial pattern recognition is the problem of selecting a 'good' set of properties to begin with . . . . If models for the pattern classes can be formulated, they can serve as a guide for selecting useful properties. (p. 158, italics added)

Put a somewhat different way, Rosenfeld says, "It is therefore of interest to measure the information content of pictures and to devise encoding schemes for representing a picture as compactly as possible."

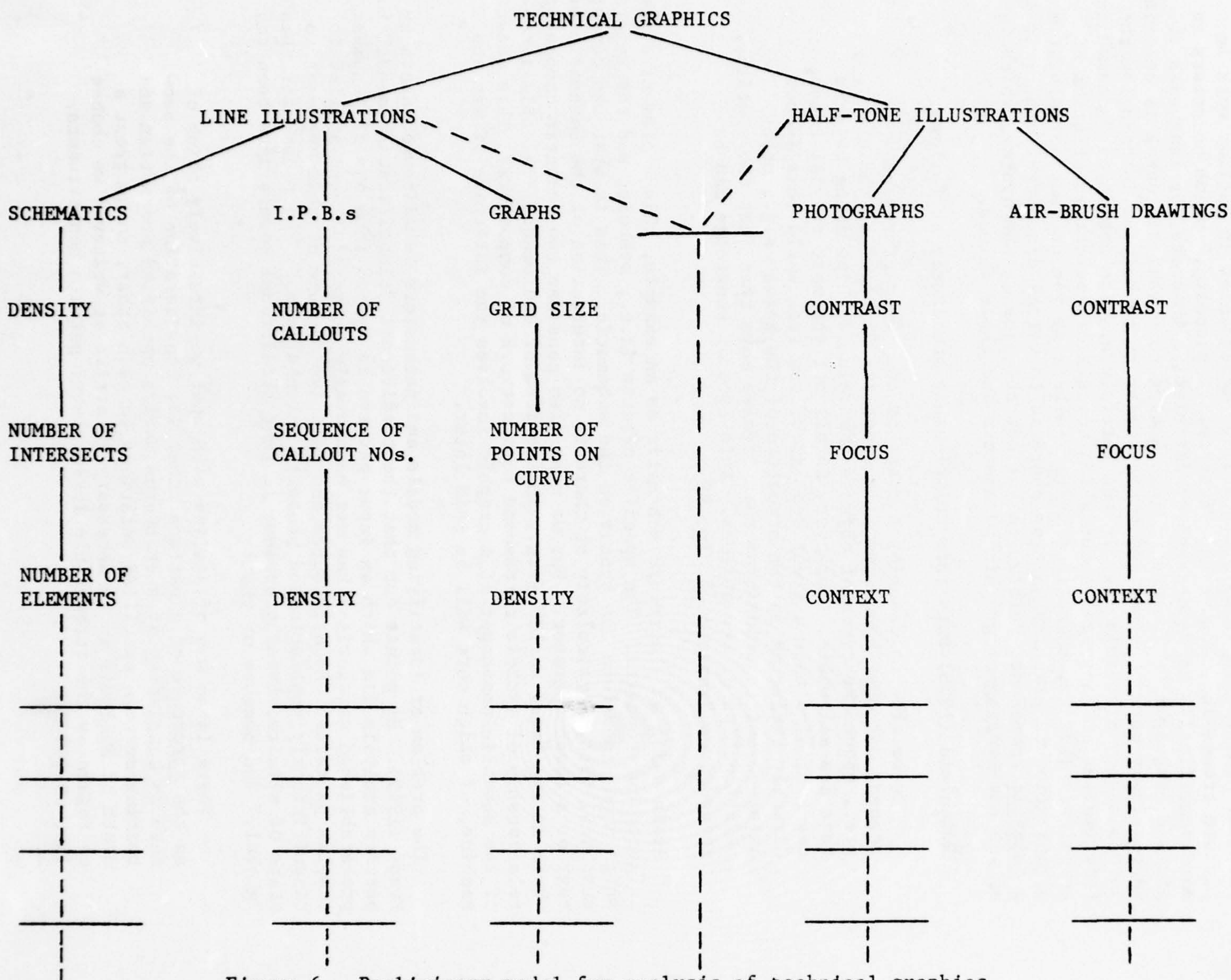

Figure 6. Preliminary model for analysis of technical graphics.

Other authors speak of such "encoding schemes" in a somewhat different fashion. Narasimhan (1969), discussing inadequacies in earlier work in picture processing, argues for "descriptive schemata," which he refers to more specifically as "syntax-based descriptive schemata, or languages . . . picture languages." The use of this analog to verbal discourse is becoming increasingly more popular, and would appear to be helpful in reducing the inconsistency of terminology in the literature. The notion of "schema" in fact implies a "language" about which to talk of the characteristics of graphics. That is, there is a syntax (albeit as yet unclear) or a systematic organization to graphics just as there is to verbal discourse. There is also a semantic component, related to but not the same as the syntax, which involves the meaning of graphic parameters and their values.

Rosenfeld (1973) puts this relationship succinctly as follows:

> To be able to describe a class of complex pictures or scenes, one must know something about the "syntax" of the class, i.e., about the types of objects and relationships among objects that are allowable. Moreover, in all but the most trivial cases, one must also know a great deal about the real world that is not directly reflected in the structure of the scene--e.g., about inter-object relationships that involve more than just the spatial relationships of the objects. This type of knowledge can be regarded as "semantic." (p. 89)

Again using an electronic schematic as an example, this statement directly reflects our concerns. The specific objects (i.e., symbols) and the relationships among them form the syntax of the schematic. This is what the computer must deal with, particularly if there is no interaction of the schematic with text or a human operator. But we must also consider the semantic component: where in the "real world" (e.g., in the item of equipment) does this interrelationship of symbols correspond to a network of components? This question of the semantic component of a graphic involves the interface of man and machine, of which more will be said later.

The problem of identifying models and parameters is addressed again by Evans (1969). He points out that the traditional "classification model" for pattern analysis (in which an input pattern is sorted into one of a number of preestablished categories) has not been overly successful when applied to complex patterns. He attributes the relative failure of this approach to "insufficiently sophisticated property calculations." The reason for this state of affairs, Evans proposes, is that traditional models have been too "global," and because of that:

> There is no way of dealing with what we intuitively think of as the structure of a pattern; that is, consideration of the pattern as consisting of subpatterns having specified properties and bearing certain specified relations to each other, to as great a depth of analysis as is necessary to arrive at whatever we choose to regard as the irreducible lowest-level pattern constituents. (pp. 80-81)

Evans here is not making a case for such relatively trivial (for our purposes) procedures for identifying the letter "A." Rather, he espouses a "model of description making," which is general enough to include a wide variety of interesting cases and is still capable of being usefully specialized to handle them. His recommendation for overcoming these present problems (similar to those made by Narasimhan & Rosenfeld) is adherence to a "syntax-directed" model for analysis of complex problems. In essence, what he is saying is that not only must a pattern recognition system be able to identify the symbols for a ground, a resistor, etc., but it must also provide us with the overall structure of the pattern--the interrelationships among the elements.

Progress to Date in Comprehensibility of Graphics

The preceding section pointed out our limited knowledge of the factors affecting graphic comprehensibility. Several sources have, however, put forth comments and/or suggestions dealing with the use of graphics in military publications. It is suspected that most, if not all, of these comments are speculative or based on intuition, and are not empirically validated. They are, however, essentially all that is available, and should be useful in alerting illustrators to characteristics of graphics which may affect their usability.

Joyce and Chenzoff (1974) point out the problem of inadequate criteria for reliable decisions about detail and size of illustrations in Job Performance Aids (JPAs). They suggest three possible reasons for deviations from "the ideal," apparently derived from close examination of existing manuals. First, they state that "contractors are tempted to make a small number of 'versatile' illustrations (i.e., detailed enough to support several task steps), rather than many specific ones." If this problem exists, it is certainly subject to experimental evaluation. Secondly, they state that "most formatters discover that, if they err by choosing an illustration size that is too small, lines run together or wash out completely, and the illustration becomes illegible. So they tend to err on the 'safe side' by choosing an illustration size that is a bit larger than necessary" (p. 11). While this may be true, it seems clear that such persons err in an appropriate direction, i.e., in a direction that improves comprehensibility. And thirdly, these authors state that "many 'versatile' or multi-purpose illustrations are eventually revealed as being sufficiently inferior to special-purpose ones, either in level of detail or angle of view, or both, so that they must be supplemented with additional 'cutaway' or 'rotated' views" (pp. 11-12). These three "variations from the ideal" alone, if shown to be valid, could be of major import in achieving higher quality graphics.

Post and Price (1974), in their style guide for technical manuals, also offer suggestions for production of comprehensible graphics. Whether or not there is an empirical base for these suggestions is not known, but intuitively they would appear to be helpful. Among the "tests" they recommend for technical writers/illustrators are several relating specifically to illustrations. These are paraphrased below in question form.

1. Is the repair cycle overview presented diagrammatically, or are the sequential relationships "buried" in prose? (Test 3)

2. Do presentations contain irrelevant information (particularly schematics and pictorials)? (Test 6)

3. Is information "compression" present (i.e., is there too much relevant information in a given image area)? (Test 7)

4. Is there an appropriate balance between narrative and pictorial material? (Test 9)

5. With regard to layout review, are all graphics referenced in the text when they are used to support a conclusion drawn in the text? Are prose and related graphics arranged for simultaneous viewing? Have small graphics been placed at the bottom of the page? (Test 17)

Probably all would agree that a writer/illustrator applying these "tests" to his product would produce a more comprehensible manual (in terms of graphics) than if he did not take these factors into consideration. It seems equally clear, however, that (for example) if a person knew (on the basis of empirical evidence) what the "appropriate balance" between narrative and graphics was, his product would be even better. Post and Price do offer some suggestions of this sort. For example, they point out that schematics should be limited to "3 elements per square inch or less," and that "important dimensions of drawings should be at least 1/10 inch." How these specific guidelines were arrived at is not explained, although they seem intuitively sound.

Kern et al. (1974), in their style guide for Army writing, list three "reasons" for using illustrations. These are: (1) to reduce detail burden in the text, (2) to emphasize points in the text, and (3) to motivate users to read the material. In discussing these points, they address the integration of text and graphics, the extension or expansion of information in text, the visual "confirmation" of information in text, and the highlighting or summarizing of main points in the text. A number of "before" and "after" examples are provided to the writer to illustrate these points. The suggestions offered in this manual, as that in Post and Price, would no doubt be of value to an illustrator/writer, but likewise lack the empirical underpinning necessary for a definitive guide for technical writers. The emphasis in Kern et al. is on providing for the user the most concise possible information on "what he will see, what he does, and how he does it," which most would agree is a necessary ingredient in technical writing.

The treatment of graphics in the two style guides discussed above (and in virtually all others as well) clearly illustrates the paucity of specific information on their effective use. Both address primarily the problem of merging text with graphics, and neither offers a significant amount of information as to the characteristics (i.e., parameters) of graphics that promote comprehensibility or usability in their own right. Of the 177 pages of their manual, Kern et al. (1974) devote only 20 pages to "using illustrations." A chapter titled "Writing and Revising--A Before and After Guide,"

by way of contrast, is approximately 135 pages in length. Post and Price (1974) do not even devote a separate chapter to illustrations--in itself an indication of how little is known about the subject. One cannot seriously fault these authors and others for slighting this problem, however. The fact is that no substantial evidence is available for specific recommendations to producers of graphics.

Both Kern et al. and Post and Price offer excellent examples of what are intuitively good and bad illustration techniques. To illustrate some extremes that actually occur in technical manuals, however, some examples are offered here. Figures 7 and 8 show an actual diagram of front panel controls for a sonar set in actual size and detail, and one page of its accompanying two and one-half page key (from NAVSHIPS 92154.2(A)). Intuitively, one would expect that this diagram in Figure 7 could be improved by increasing it to full page and/or reducing the number of call-outs (as well as by other changes). It is believed that Post and Price would include such an example in their concept of "information compression." In this instance, the diagram was accompanied by another very simple (and related) diagram and one-half page of text. It was apparently more important, in the eyes of this writer/illustrator, to present the diagram and accompanying text and figure on the same page than to change it as suggested above. Figure 9 shows another diagram from the same manual, in actual size and detail. The size of significant features would seem to indicate that this diagram could easily occupy a much smaller space. All references to these diagrams are based on the intuition of this writer, with no empirical evidence as to their validity. It seems, in the area of graphics, that this is the only recourse at the moment. Research and development in the area are clearly needed.

The Navy Personnel Research and Development Center is currently conducting an exploratory investigation in the field of graphics. A major goal of this endeavor is to develop a model or schema from which, hopefully, parameters associated with quality of graphics can be derived. A more distant goal of the investigation is to develop a measure of "graphic worth" which is objective in the sense that individual graphics need not be "judged" by human observers (a very costly and time-consuming process). Progress to date on this study is limited. An initial bibliographic reference file has been established on visual information processing and related psychological matters. One area which was considered to offer some promise in this work was that of Photo Interpretation (PI), as carried out primarily by the intelligence community. Exploration in this area, however, has not lived up to expectations. As with other facets of graphics, specific parameters of photos are understood only in gross terms, and the major strategy of PI is intuitive. It is considered an art more than a science. One parameter which has been clarified for pattern recognition in general is that of "contrast." Considering that the cone of foveal vision (within which most visual information is acquired) subtends an angle of only 1.5 degrees, resolution within this cone can be examined. Resolution is defined here as the "equivalent number of points (pixels) that can be separately identified." Resolution itself is a function of the brightness range. With greater light intensity there is greater resolution up to the limit where greater brightness becomes distasteful and even harmful. Most simply, there are differences of level of brightness on a gray scale. Ordinary vision can separately distinguish many levels of gray scale in the usual range of brightness. It remains to be determined, however, what the optimum gray scale range is for various kinds of graphics.

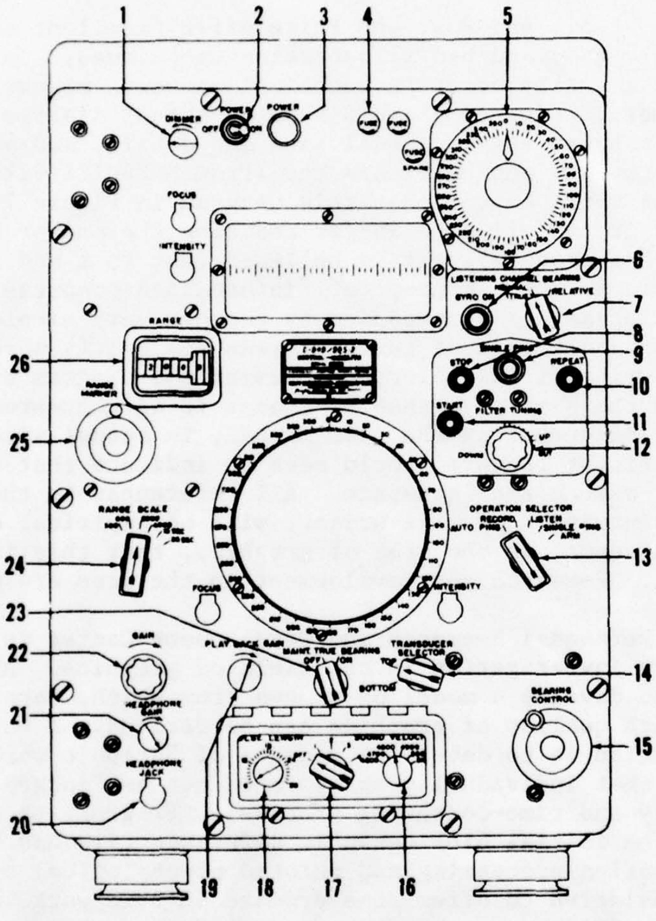

Figure 7. Example of possible information "compression" in technical graphics.

KEY TO FIGURES 4-1, 4-2, 4-3 and 4-4		
Index No.	Nomenclature	Function
1	DIMMER Knob	Controls the brilliance of the front panel illuminating lamps for RANGE counter and the LISTENING CHANNEL BEARING indicator.
2	POWER Switch	Applies or removes power to the sonar equipment in the ON or OFF positions respectively. Controls the primary power for all units of the system with the exception of 115 volt ac service outlets and the electronic rotor temperature control oven.
3	POWER Pilot Lamp	Illuminates to indicate that primary power is applied to the sonar system.
4	FUSE	The two fuses are employed in the primary power circuits. The third fuse holder contains a spare fuse.
5	LISTENING CHANNEL BEARING Indicator	Indicates azimuth bearing to which the listening system is trained. The indicator is synchronized with the servo system and provides an indication of the transmission bearing when single ping operation is employed.
6	GYRO ON Lamp	Illuminates when the submarine gyro voltage is applied to the sonar system. The lamp must be illuminated before true bearings can be obtained.
7	NORMAL (TRUE)/ RELATIVE Switch	Permits use of the ship's gyro excitation voltage in the sonar servo system. The bearing indications of both the PPI and LISTENING CHANNEL BEARING indicator are true or relative as selected by this switch.
8	STOP Button	Halts the action of the magnetic recorder during the repetition cycle.
9	SINGLE PING Lamp	Illuminates when single ping system is ready for transmission. Lamp is extinguished when the START button is depressed and the single ping cycle is started.
10	REPEAT Button	Initiates the repetition function of the magnetic recorder when single ping operation is employed.
11	START Button	Initiates the transmission and record functions of the magnetic recorder when single ping operation is employed.
12	FILTER TUNING Knob	Controls circuits which determine the frequency of echo signals except during repeat operation for display on A-scan indicator. This control is used for single ping operation only.
13	OPERATION SELECTOR Switch	Selects mode of operation to be performed by the sonar system. The switch is provided with five positions. The fifth position is spring-loaded so as to return to the fourth position. The five positions are as follows:
a.	AUTO PING Position	Automatic transmission, reception and display as described in paragraph 1a of this section.

Figure 8. Partial key to control-indicator illustration in Figure 7.

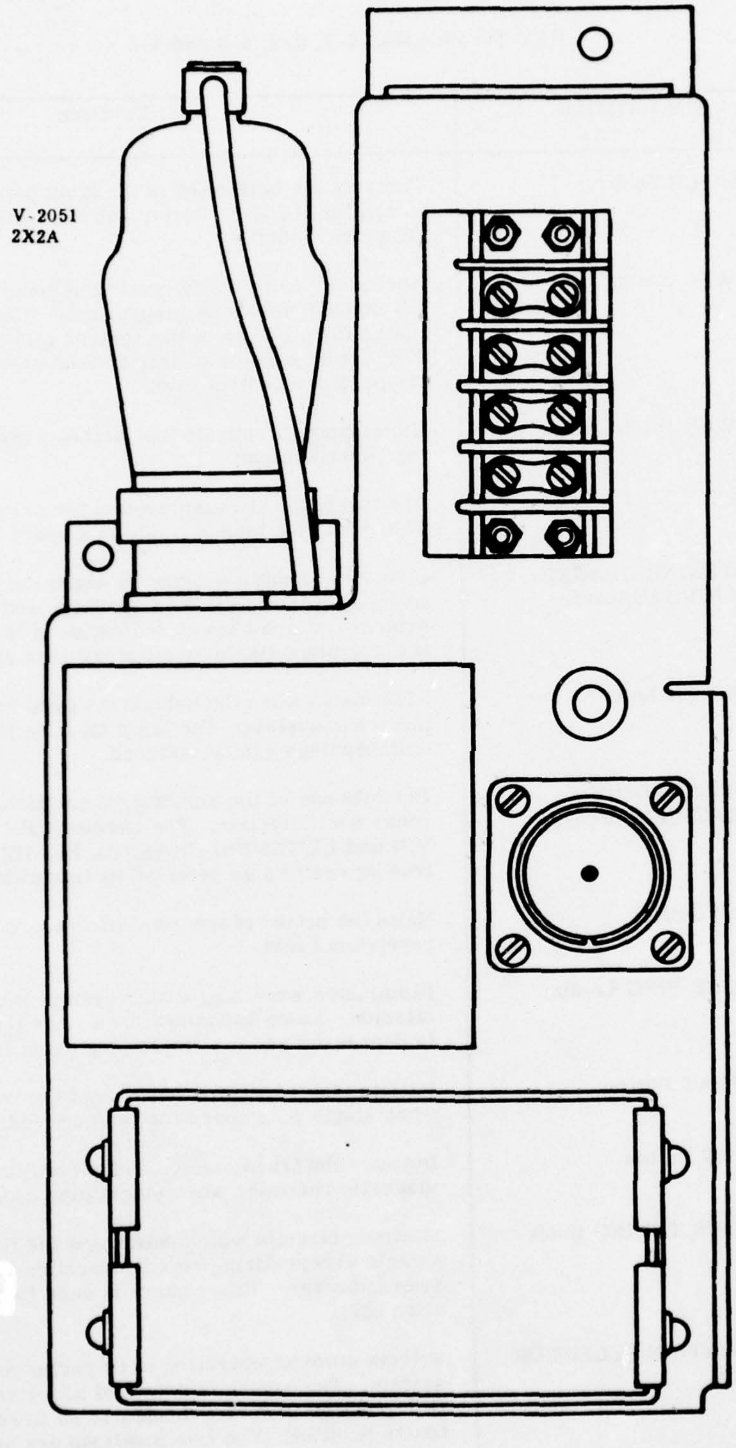

Figure 9. Example of possible overexpansion of technical graphic.

A second direction taken in the NAVPERSRANDCEN investigation is an examination of the literature dealing with "aesthetic values" in graphics. Variables said to be involved in aesthetics include symmetry, variety, uniformity, simplicity, intricacy, quantity, and "convincing representation." It is not known at present how these variables apply to graphics in general, but it is suspected that they do have some relationship to graphic comprehensibility. The interested reader is directed to the work Aesthetic Measures, by George D. Birkhoff (1933), for a full treatment of this area.

In summary, the NAVPERSRANDCEN investigation has thus far resulted in few "hard" findings; in fact, one might say that it has resulted to date in more questions than answers. This work will be continued for a brief time, in the hope that at least a preliminary model of pattern recognition can be established.

Before proceeding to a discussion of the state-of-the-art of automation of graphic assessment/production, a word should be said of the potential contribution of psychology to the area of graphic parameters. Perception and memory would seem to offer the most in this regard. Briefly, perception is not, as sometimes thought, a passive internalization of the outside world. Rather, it is ". . . an active process of synthesizing or constructing a visual figure" (Neisser, 1967). As Neisser further points out: ". . . such a complex constructive act must take a certain amount of time." It is this interval of time that is important to the present discussion. It is difficult to say exactly when perception stops and memory takes over. If one examines a graph (for example) for a very brief interval and then looks away, the process of perception continues. Although it is not precisely known how the process operates, it is clear that, even after the stimulus is removed, a "visual image" (in some sense) remains. A person is able to perform operations on what he has seen during the brief time that the image is "stored" in what many psychologists term "iconic store," or "iconic memory." Although research in this area has involved relatively simple stimuli, there is every reason to believe that the same phenomenon applies to complex ones such as graphics. The perceptual process (including "iconic memory") must be distinguished from that which we commonly refer to as "memory." Obviously, we "remember" things for much longer periods of time than the few seconds of "iconic memory." Feigenbaum (1970) presents one of the more popular theories of memory. He hypothesizes that there are three types of information storage structures: an immediate memory, an acquisition memory, and a permanent store. (Other terms one might encounter for these three structures are "sensory register," "working memory" or "short-term memory," and "long-term memory.") Immediate memory may or may not be identical to iconic storage. It is possible that this structure extends beyond the few seconds of perceptual processing. It does, in any event, seem to be of import with regard to on-the-job use of graphics. Of this structure, Feigenbaum says:

> It is a buffer storage mechanism of extremely small size, holding a few symbols. Inputs from the peripheral sensing and encoding mechanisms are held here in a state of availability for further central processing . . . . The net result of such an immediate memory mechanism is that the total processing system has a very narrow "focus of attention," that is, the central processes can attend to only a miniscule portion of the environment at any time. (p. 455)

Whether this immediate memory and iconic memory are the same or not, it is at least intuitively obvious that there are parameter values in graphics that will enhance this short-term storage. These should be taken into consideration in graphics research. Further, since we would like to think that technicians not only use TMs but, in addition, learn from them, an investigation of optimum graphic parameters should necessarily include a consideration of the overall three-level memory structure.

Automation of Graphic Assessment and Production

As with text, the extremely large volume of materials with which we must deal argues strongly for the highest possible degree of automation. For the purposes of this paper, only a brief overview of the state of the art of computer graphics will be offered. As indicated in the foregoing sections, so little is known of the parameters contributing to graphics comprehensibility that virtually no attention has been given to such parameters in the context of production automation. Suffice to say that the art of graphics automation has so far outstripped our understanding of what to automate.

The general nature of the potential of computer graphics is stated clearly by Parslow, Prowse, and Green (1968) when they say:

> The graphics terminal opens up a completely new range of fields of application for computers . . . . The potential recently revealed for interaction between computer and user is vitally important for the greater application of these machines in all spheres of industry, commerce, and scientific development. The interactive graphic terminal transfers the computer from a cumbersome specialist "tool" into a "colleague" helping to work out a solution to problems during a dialogue through the common visual link. "One picture is worth a thousand words," and in computing it is certainly true that one picture can be considerably more valuable than several yards of lineprinter output. This is all the more true if a scientist or a business executive has to interpret the output and take further action on it with the computer. (p. ix)

In essence, one can think of the area of computer graphics as a problem of communication between man and machine. As in any automatic data processing (ADP) problem, the need is to tell the machine what it is we want it to do. Because of our relative lack of knowledge about what this is (due to our naivete about important parameters of graphics), the ideal communication system must be a fast-reacting, give-and-take interchange; e.g., the interactive graphics console. A critical factor is the language for this interchange. Narasimhan (1969) puts this problem as follows:

> In the last few years, with the increasing availability of interactive consoles and computer-aided graphic design systems, the generative and descriptive aspects of picture processing have assumed central importance. In the context of man-machine interactive systems dealing with visually given data, the need is for two languages: (1) a picture language . . . and (2) a textual, conversational or discourse language . . . [to handle the dialogue between man and machine]. (p. 5)

Narasimhan is quick to point out that only rudimentary aspects of the problems associated with these two types of languages have so far been addressed. He says that an adaptive, interactive picture processor having these language capabilities does not yet exist. He speculates that the major reason for this lack concerns the languages themselves. Narasimhan maintains that the problem of graphic language cannot be meaningfully addressed until one has "some operationally viable model of language and language behavior," which the field of psychology does not yet have.

The development of computer techniques to facilitate graphic processing has greatly outstripped the understanding of psychological factors such as the language or syntax of graphics. No attempt will be made to describe in depth the development of such computer hardware. Suffice to say that techniques such as the following have greatly enhanced our ability to deal with graphic forms.

1. Time-sharing: makes it economically feasible to give a single user access to a large computer on an immediate local basis with quick response time and fast turnaround.

2. Graphic Data Processing: allows the user to communicate with the computer conveniently and in his own terms. It allows the computer to communicate information to the user in a form which is compact and descriptive without the intervention of a programming language.

3. The Light Pen: a light sensor which, when placed adjacent to a given area on a cathode ray tube, indicates to the computer where it is pointing and that some action is to take place at that point.

4. The Data Tablet: a flat surface with interface to a computer and a stylus, the position of which can be detected on the tablet in terms of X-Y coordinates. An operator can start with hard copy, source data, photographics projections, or what have you and input these to the computer simply by tracing them with the stylus.

5. Optical Character Recognition: an input system with the capability of digesting information at a rapid rate, converting text or graphics directly to a form suitable for computer processing, without intermediate conversion to coded media.

The Navy currently has, in developmental stages, two systems which are intended to examine these technological advances in terms of potential benefit to the Navy. The first of these, OLPARS (On-Line Pattern Analysis and Recognition System) is undergoing test and evaluation at the Naval Ship Research and Development Center, Carderock, Maryland. This is an interactive graphics system which will be used to explore the structure of n-dimensional data via a variety of projections onto the two-dimensional CRT. Its primary use as of now is in pattern analysis (notably "cluster analysis"), rather than in comprehensibility, but the potential is clearly present for research and development into the latter, more basic problem. As a first step, the system might be used to generate families of illustrations that could then be submitted to empirical tests of comprehensibility, usefulness, etc. Hopefully, this would yield a better understanding of parameters relevant to these factors, after which the system could be adapted to interactive <u>production</u> of illustrations incorporating these parameters.

A second system being examined by the Navy is TRUMP (Technical Review and Update of Manuals and Publications), situated at the Naval Air Rework Facility, Jacksonville, Florida. A major capability of TRUMP is Optical Character Recognition. The Naval Air Systems Command (1974) points out that a typical TRUMP configuration has input capacity challenging the rate of 40 to 50 input keyboard operators. GRAFIX I, the core computer data entry system for TRUMP, can potentially handle all printed, typewritten, and hand-lettered material. In addition, it has the potential for digitizing tables and illustrations, although the problems of the vast storage requirement must be solved before this can be done efficiently. The system does have the capability to draw graphics (which are stored photographically) on a CRT, and thence record them on photosensitive paper. The TRUMP system clearly has the potential for a wide array of techniques for producing Navy documentation. Little concern has been paid thus far to the psychological factors in graphic comprehensibility, but as new knowledge is achieved in this area, TRUMP can potentially put it to good advantage.

In summary, advances are being made at a rapid rate in the field of automated graphic processing, but factors related to comprehensibility of graphics have not been serious considerations in such advances. The state of the art in the field is such, however, that as new facts become available existing systems can easily take them into account.

CONCLUSIONS AND RECOMMENDATIONS

Based on the findings in the areas of readability and comprehensibility of both text and graphics--in terms of previous research, theoretical discussions, and the current state of the art--the following conclusions and recommendations are put forth.

1. Comprehension Variables. Relatively little is known regarding those factors or variables in writing that have a causal relationship to comprehension. In other words, we have virtually no basis on which to specify for a writer those manipulations which he can make in his work which will tend to make it easier to understand. It is recommended, therefore, that research be instituted to identify such variables, and that these be expounded in such a way that they can be used by writers of Navy materials. (There follows an extensive annotated bibliography of previous research in this area, which, while no doubt of value, is *in toto* inconclusive.) This research program must take into account the personnel characteristics of the user or consumer and the environments in which the material is to be used. It must address such factors as motivation of the users, the tradeoff between accuracy and speed of reading and comprehension, and the relationship between "acceptability" and readability level of various types of technical writing. This latter question refers to the fact that writing might be made excessively simple in order to ensure comprehension, but in the process result in a feeling by the reader that he is being "talked down to," and therefore not be accepted readily.

2. The TRUMP System. The TRUMP System (Technical Review and Update of Manuals and Publications) is a technical manual processing system currently undergoing test and evaluation at the Naval Air Rework Facility, Jacksonville, Florida. (A detailed description of the working of the system is given in the Appendix.) Considering the tremendous volume of written material comprising the Navy's technical documentation system, it seems clear that a central and highly flexible computerized system is required. It is recommended that TRUMP be considered for use as the core for such a capability. While its present primary function is the revision and storage of technical data, it would appear to have potential for many other aspects of technical manual processing. Development of on-line interactive capability, sophisticated readability analysis (using a technical dictionary, key-word-in-context storage and retrieval, etc.), and remote time-sharing are recommended for future inclusion in the system, based on the results of interim research to be described below.

3. Optimum Readability Formulas. As pointed out earlier, there are a number of readability formulas in existence which could be considered as candidates for assessment of difficulty of technical writing. But before any formula can be said to be "best" for any particular purpose, several preliminary questions must be addressed. The first of these issues is to determine the degree to which technical terms "inflate" the actual difficulty level of material when measured by formula. It is recommended that research be undertaken first to identify those technical terms which

would potentially be problematic at various levels of expertise of technicians. If these terms are indeed found to unduly influence the application of formulas, alternative methods for accounting for such terms, such as use of a dictionary, counting selected words as "easy" (i.e., having no more effect on comprehension than one- or two-syllable common words), counting such terms only the first time they occur in a sample, or even not counting such words at all, should be developed and tested. A second area where readability formulas should be examined is in the initial writing stages. Using the information from the first research effort, it is recommended that at least two readability formulas (which can be computed either manually or by computer) be specified for technical writers. Having such formulas, writers can assess the difficulty of their material as they proceed in the early drafts and editors can check (or cross-check) technical writing at appropriate points in the procedure. The parameters for initial test of these formulas should be sentence length and word length (based either on a syllable count or use of a word list). The readability of end products with and without the application of such formulas should then be carefully researched.

A third research question with regard to readability formulas is to determine the variables which, in combination, best predict the difficulty of technical writing. The result of this approach would be the identification of (probably) a very complex formula, containing (perhaps) many more than two variables, and quite likely necessitating automated computation. The writer should be made aware that this formula will be the ultimate test of the difficluty of his writing, but the complexity of the formula should prohibit the writer from simply minimizing difficulty on each of its variables (i.e. "writing to formula").

4. Guiding Technical Writers. Many indications point to the fact that writers of Navy materials fall into one of three categories: subject matter experts who have little or no training or expertise in writing, technical writers who have little or no background in the intricacies of hardware design and construction, or persons who are neither subject matter experts nor good writers. We believe it only fair to say, of course, that there are those in the fourth category who can write and do understand the subject matter. It is considered vital to the optimization of technical manuals that all writers be moderately well versed in the subject matter and have a very good understanding of the writing craft. It is recommended, therefore, that a research and development program be undertaken to determine and implement an effective procedure by which technical writers can be aided in improving their products. Initially TM material should be reformatted using one or more existing style guides, and comprehension of representative users compared on the original and revised versions. Any available input from the results of recommendation 1 above (identifying variables causally related to comprehension) should be used to supplement the style guides used. Based on the results of the comparison of original and revised versions of a manual, alternative forms of style guide(s), and/or a new style guide comprising the best components of existing ones should be developed. The new guide(s) should then be used to again reformat TM material (with comparisons as above) and the results of this second iteration compared to the initial study. The readability, comprehensibility, and usability of the style guides themselves (including their "acceptability" to the

writer) should be examined during the course of the program. If the original rewriting was not accomplished by actual technical writers, a final version of the style guide should include this step. The project should also ideally include a phase which examines whether or not formal training in writing and/or the use of the style guide improves the readability, comprehensibility, and usability of the final product. If it is found that a significant improvement in writing can be achieved through this training, a "technical writer course" should be developed for Navy use with its contracting agencies as well as for in-house writing. The question of whether a "screening device" for prospective writers would be of value should also be considered in this latter phase.

5. Computerized Feedback to Writers. Recommendation 3 had to do with providing writers a formula by which they could check their writing at a very early stage. Cost-effective time-sharing readability analyses are available which would permit readability checks plus rewrite assistance as the writer reaches a more smooth form of his manuscript. (A detailed description of one such system--that of Scientific Time Sharing, Inc.-- is given in the Appendix.) It is recommended that contracting agencies as well as Navy writers be given access to such a capability on an experimental basis. Having access to such a system, a writer may take a sample of his work to a conveniently located desk-top console, type it in, and receive immediate information as to its readability characteristics. Research and development are required to determine what further potential such a system has in terms of storing a technical dictionary and providing the writer with alternatives to difficult words which may occur in his material. There seems to be great promise of such potential. A future complement to such a system might very well be an index of various syntactic constructions which result in more or less difficulty in comprehension. This latter capability is believed to be presently beyond the state of the art, but could potentially be added to such a system at a later date.

6. An Integrated Technical Manual Production System. Proceeding in a stepwise progression, it is considered unwise to jump to a complete system which one assumes will solve all technical documentation problems. Intermediate research and development steps are advised, including the ones outlined above. CARET I (the computerized system developed by Klare, et al) offers a point of departure from which to try new ideas, based on outcomes of earlier research steps. It should be able, with modification, to provide detailed readability analyses, editorially composed printout, ready-to-administer "cloze test," feedback to writers dealing with alternative words, phrases, or sentences which should be more comprehensible, and a "lexicon" of technical terminology to which the writer might attend upon beginning his writing task (e.g., a list showing the frequency, association value, etc. of words). This program in a basic form is already available, only amplification is required. It should be noted that the recommendation for an integrated system overlaps many of the other recommended steps above. The key word here is "integrated," and by the very nature of its underlying rationale would incorporate the outcomes of the various other lines of research.

7. Parameters of Comprehensible Graphics. As seen earlier, little is known of the variables making up illustrations of various kinds and their relationship to the comprehensibility of that illustration. The guidance which is offered technical illustrators is for the most part intuitive; virtually no empirical evidence is available. It is recommended that research be continued with respect to identification of such parameters. Once the basic parameters have been specified and a schema developed for guiding illustrators, specific guidelines should be incorporated into both style guides and military specifications to assure their use in technical documentation production.

8. Automated Production of Graphics. The state of the art of computers is such that, in the area of graphics, an engineer can make up a rough draft of a schematic, flow chart, etc., have it scanned and stored by computer, present it on an interactive terminal, provide the operator with "cues" and software capability to optimize it, photograph the result, and print it in copyready form. American Telephone and Telegraph Company is currently using such a procedure, as are, no doubt, other commercial enterprises. This author has observed the output of such a system and would compare it favorably with precise, intricately drawn illustrations from drawing boards, printed through standard procedures. The Navy has available OLPARS (On-Line Pattern Analysis and Recognition System) which could be a very valuable vehicle for systematic research and development in the area of graphic design and production. A description of the OLPARS is given in the Appendix.

REFERENCES

American Bible Society. Good news for the modern man: The New Testament in today's English version (3rd Ed.). New York: Author, 1971.

American Institutes for Research in the Behavioral Sciences. Word frequencies in technical instruction. Prepared for Personnel Research Division, Bureau of Naval Personnel, Department of the Navy, Contract NOp-1498, August 1964.

American Telephone and Telegraph Company. A new approach to maintenance BSPS: Task-oriented practices--Plant (TOPP). New York, Author, 1974.

Bates, E. S. (Ed.). The Bible: Designed to be read as living literature. New York: Simon & Schuster, 1951.

Biersner, R. J. Reading grade levels of Navy rate training manuals and non-resident career courses. (CNETS Report 2-TS). Pensacola: Chief of Naval Education and Training Support, May 1975.

Birkhoff, G. D. Aesthetic measure. Cambridge: Harvard University Press, 1933.

Bormuth, J. P. Development of readability analyses (Final Report, Project No. 7-0052). Washington, D.C.: U.S. Department of Health, Education, and Welfare, USOE, Bureau of Research, March 1969.

Bransford, J. D., & Johnson, M. K. Considerations of some problems of comprehension. In Visual information processing, W. G. Chase (Ed.). New York: Academic Press, 1973, 383-438.

Brown, W. R. Science textbook selection and the Dale-Chall formula. School Science and Mathematics, February 1965, 65, 164-167.

Carroll, L. Through the looking glass. New York: E. P. Dutton & Co. Inc., 1954.

Carver, R. P. Improving reading comprehension: Measuring readability (Final Report). American Institutes for Research for Office of Naval Research, May 1974.

Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. Methodologies for determining reading requirements of Military Occupational Specialties (HumRRO Tech. Rep. 73-5). Monterey, CA: Human Resources Research Organization, March 1973.

Dale, E., & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20 and 37-54.

Duffy, T. M., Carter, J. D., Fletcher, J. D., & Aiken, E. G. Language skills: A prospectus for the Naval Service (NPRDC Special Rep. 76-3). San Diego: Navy Personnel Research and Development Center, October 1975.

Evans, T. G. Descriptive pattern analysis techniques. In A. Grasselli (Ed.), Automatic interpretation and classification of images. New York: Academic Press, 1969, 79-95.

Feigenbaum, E. A. Information processing and memory. In D. A. Norman (Ed.). Models of human memory. New York: Academic Press, 1970, 451-468.

Fetter, W. A. Computer graphics in communication. New York: McGraw Hill, 1965.

Flesch, R. F. Measuring the level of abstraction. Journal of Applied Psychology, 1950, 34, 384-390.

Flesch, R. F. A new readability formula. Journal of Applied Psychology, 1948, 32 221-233.

Fry, E. B. A readability formula that saves time. Journal of Reading, 1968, 11, 513-516.

Gunning, R. The technique of clear writing. New York: McGraw Hill, 1968.

Hamilton, E. A. Graphic design for the computer age. New York: Van Nostrand Reinhold Co., 1970.

Harris, A. J., & Jacobson, M. D. Basic elementary reading vocabularies. New York: Macmillan Publishing Co., 1972.

Harris, A. J., & Jacobson, M. D. Some comparison between basic elementary reading vocabularies and other word lists. Reading Research Quarterly, December 1974, 87-109.

Harris, A. J., & Jacobson, M. D. Comparative predictions of tested reading comprehension of intermediate and secondary pupils. Paper presented at Annual Convention of the American Educational Research Association, Washington, D. C., April 1975.

Horn, S. Light pen input to a design package in current production. In R. D. Murray (Ed.). Computer handling of graphical information. Washington, D. C.: Society of Photographic Scientists and Engineers, 1970, 41-52.

Huggins, W. H., & Entwisle, D. R. Iconic communication: An annotated bibliography. Baltimore: Johns Hopkins University Press, 1974.

Johnson, K. H., Relova, R. P., & Stafford, J. P. An analysis of the relationship between readability of Air Force procedural manuals and discrepancies involving non-compliance with the procedures. Unpublished Master's Thesis, Air Force Institute of Technology, Air University, August, 1972. (Document No. AD 750 917, National Technical Information Service.)

Joyce, R. P., & Chenzoff, A. P. Improving job performance aids through condensation, dual-level presentation, promotion of learning, and entry by malfunction symptoms (AFHRL-TR-74-12). Wright-Patterson AFB, OH: Air Force Human Resources Laboratory, March 1974.

Kern, R. P., Sticht, T. G., Welty, D., & Hauke, R. N. Guidebook for the development of Army training literature (Tech. Rep.). Arlington: U.S. Army Research Institute for the Behavioral and Social Sciences, 1974.

Kincaid, J. P. Making technical writing readable (the Fog Count and an alternative). Human Factors Society Bulletin, 1972, 15(5), 3-4).

Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel (CNTT Res. Rep. 8-75). Memphis: Chief of Naval Technical Training, February 1975.

Klare, G. R. The measurement of readability. Ames, IA: Iowa State University Press, 1963.

Klare, G. R. Assessing readability. Reading Research Quarterly, Vol. X, No. 1, 1974-1975, 62-102.

Klare, G. R. A manual for readable writing. Glen Burnie, MD: REM Company, 1975.

Klare, G. R., Rowe, P. P., St. John, M. G., & Stolurow, L. M. A first version of a computer-aided revision, editing, and translating system (CARET I) (Tech. Rep. No. b.2.1). Cambridge: Harvard University Computer Aided Instruction Laboratory, 1969.

Kulp, M. J. The effects of position practice readability level on performance. Unpublished Master's Thesis, San Jose: San Jose State University, 1974.

Machover, C. Computer graphics in the United States. In R. D. Parslow, R. W. Prowse, & R. E. Green (Eds.). Computer graphics: Techniques and applications. New York: Plenum Press, 1969, 61-83.

Matsa, S. M. What has computer graphics to offer? In R. D. Parslow, R. W. Prowse, & R. E. Green (Eds.). Computer graphics: Techniques and applications. New York: Plenum Press, 1969, 1-6.

Maxwell, J. A. Hobo symbology. In E. Whitney (Ed.). Symbology: The use of symbols in visual communications. New York: Hastings House, 1960.

McElroy, J. The Fog Count. In U.S. Air Force Manual 11-3. Maxwell, AL: Maxwell Air Force Base, Air University, June 1953.

MIL-M-24100B. Military Specification: Manuals, Technical, Functionally Oriented Maintenance Manuals (FOMM) for Equipment and Systems. January 1974.

Murray, R. D. (Ed.). Computer handling of graphical information. Washington, D. C.: Society of Photographic Scientists and Engineers, 1970.

Narasimhan, R. On the description, generation, and recognition of classes of pictures. In A. Grasselli (Ed.). Automatic interpretation and classification of images, 1969, 1-42.

Naval Air Systems Command, Technical Data Management Projects. MIARS and TRUMP mechanized update techniques. Author, 1974.

NAVSHIPS 92154.2(A). Operator's handbook for sonar set AN/BQS-2. 1954.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Parslow, R. D., Prowse, R. W., & Green, R. E. (Eds.). Computer graphics: Techniques and applications. New York: Plenum Press, 1969.

Post, T. J., & Price, H. E. Requirements and criteria for improving reading comprehension of technical manuals (Final Rep.). Washington, D.C.: Naval Sea Systems Command, Technical Publications Branch, 1974.

Rosenfeld, A. Progress in picture processing: 1969-71 Computing Surveys, 1973, 5(2), 81-107.

Simons, H. D. Reading comprehension: The need for a new perspective. Reading Research Quarterly, 1971, 4(3), 338-363.

Smith, E. A., & Senter, R. J. Automated readability index (AMRL-TR-66-22). Wright-Patterson AFB, OH: Aerospace Medical Division, 1967.

Sticht, T. G. Reading for working: A functional literacy anthology. Alexandria: Human Resources Research Organization, 1975.

Stocker, L. P. A word list to be employed with the Dale-Chall readability formula for the appraisal of readability levels of Catholic materials. Unpublished doctoral dissertation, Fordham University, 1967. (Dissertation Abstracts, 1968, 28, 4496A).

Stocker, L. P. Increasing the precision of the Dale-Chall readability formula. Reading Improvement, 1971-1972, 8, 87-89.

Thorndike, E. L. Educational psychology. New York: Lemcke & Buechner, 1917.

Thorndike, E. L., & Large, I. The teacher's word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

U. S. Air Force. Guide for Air Force writing. AFP10-1, Department of the Air Force, 1969.

Whalen, G. V. Study for investigation of the application of the readability ease assessment device to Navy technical publications. St. Louis: McDonnel Douglas Corporation, 1974.

Whetstone, A. Free-hand data input. In R. D. Murray (Ed.), Computer handling of graphical information. Washington, D. C.: Society of Photographic Scientists and Engineers, 1970.

Whitney, E. (Ed.). Symbology: The use of symbols in visual communications. New York: Hastings House, 1960.

REFERENCE NOTES

1. Scientific Time Sharing, Incorporated. Press Release dated March 3, 1975.

2. United Press International. Press Release dated February 17, 1976.

3. Associated Press. Press Release, date unknown.

4. George R. Klare, Personal Communication. Further information in this regard may be solicited by writing to Dr. George R. Klare, Department of Psychology, Ohio University, Athens, Ohio 45701.

ANNOTATED BIBLIOGRAPHY:
UNPUBLISHED MANUSCRIPTS, RESEARCH REPORTS, AND JOURNAL ARTICLES
EXAMINING COMPREHENSION VARIABLES

ANNOTATED BIBLIOGRAPHY: UNPUBLISHED MANUSCRIPTS, RESEARCH REPORTS, AND JOURNAL ARTICLES EXAMINING COMPREHENSION VARIABLES

Allbaugh, N. J. Comprehension of three levels of social studies material as designated by a readability formula. Unpublished doctoral dissertation. The University of Iowa, 1968.

Selected passages of approximately 300 words at each of three corrected Dale-Chall grade levels, 4th and below, 5th-6th, and 7th-8th. Passages also differed in Dolch "fact burden" from high, to medium, to low, making a total of nine experimental passages. Prepared five four-choice items on each passage, and ordered into three test booklets. Administered in regular classrooms by regular teachers on three successive days, giving 20, 25, and 30 minutes for reading and then 20, 25, and 30 minutes for testing on the three booklets. Selected 100 Ss randomly from classes for analysis (50 M, 50 F) and compared effects of difficulty level, fact level, grade level, IQ level, and sex. Found that overall performance on test decreased as difficulty increased, as predicted by Dale-Chall formula. However, found reversals when grade levels of students were considered. Analysis of effect of fact burden found inconsistent results, with Ss performing better on medium fact level than other levels, generally. Possible contributors to mixed effects of difficulty (readability): material used (content not held constant); tests used (relative difficulty of tests on the different passages was not equated); administration (tested in regular classrooms by regular teachers, tending to yield a high and effective level of motivation).

Blue, L. L. A study of the influence of certain factors in science materials on the reading comprehension of seventh grade pupils. Unpublished doctoral dissertation, Indiana University, 1964.

Prepared 8 versions of approximately 900 words (863 to 916) on topic of Kepler's three laws of planetary motion, varying versions in Dale-Chall readability (7-8 grade level versus 11-12 grade level), style of writing (expository versus narrative), and use of author's definitions (presence or absence). Varied both non-Dale words and sentence length to achieve readability differences, but carefully kept organization and development of ideas constant. Administered text of 25 four-choice items to eight randomly selected groups of 30 seventh grade Ss, handing selection to each on entering regular class and allowing 25 minutes for reading text and taking test (also administered science survey and mental maturity tests). All versions, using readability, style of writing, and definitions, not significantly different. Possible contributors to negative results: tests used (low sensitivity shown by high means, 18 or 19, and highest scores, 24 or 25, in each version), very high and effective motivational level (tests given in regular classes with regular instructors helping, with tests handed out as class assignment with liberal reading and testing time and exhortation to finish, with test available during testing, and with topic judged high in interest). Interesting added findings: suggestion of probable larger differences raised by results of small pilot study in different schools under less motivating conditions; also, Ss judged 11-12 version harder than 7-8 in comparative part of study. Note also that Ss in comparative part who had hard version, then easy, gained more between administrations than those who had easy version, then hard; this follows a "set to learn" prediction. Carefully executed study.

Brewer, R. K. The effect of syntactic complexity on readability. Unpublished doctoral dissertation, University of Wisconsin, 1972.

Related syntactic complexity to readability, using an automated syntactic analysis program on 50 of the 390 passages in the 1950 version of the McCall-Crabbs Standard Test Lessons in Reading. Program output yielded 5 measures: (1) McCall-Crabbs comprehension scores in terms of average grade levels (for 50% comprehension level); (2) syntactic complexity measure, with number of components found at last level of parsing divided by number of sentences in passages; (3) mean number of words per sentence; (4) total number of words per passage; and (5) normalized syntactic complexity measure, with above syntactic complexity measure divided by number of words in passage. Found intercorrelations of above measures ranging from .32 to .93, with all but one of the 10 correlations significantly different from zero. Results termed disappointing because the simpler measures of words per sentence and words per passage were better predictors of McCall-Crabbs scores than the more elaborate syntactic measures. A multiple regression yielded an acceptable level of predictive power (.48), but an unacceptable level of error with the syntactic complexity measure and the words per passage. States that future automatic syntactic analysis is highly feasible.

Cervone, E. V. Achievement in and attitude towards senior high school U. S. history with reduced readability texts. UUnpublished doctoral dissertation, Temple University, 1974.

Chose two textbooks in U. S. history, one at 5-6 grade level and one at 9-10 grade level by Dale-Chall readability formula, and 4-5 and at 11 grade levels by Fry readability graph. Used "easy" text with 20 lowest-achieving students and "hard" text with 11 lowest-achieving students ("hard" text also used in regular instruction) in senior high school U. S. history classes over a three-semester period. Two dependent measures were used at three widely spaced times during course; these were: the Cooperative Social Studies Tests, American History (to measure achievement), and an attitude survey, with general questions plus one specific question on attitude toward readings and attitude toward U. S. history course. Results indicated no significant differences in Coop (achievement) scores, but significant differences in attitude, favoring those who had easy text, toward the U. S. history course and also the readings in the course. Possible contributors to mixed results: material used (though all content covered U. S. history, exact equivalence of content could not be ascertained since texts differed); tests used (though test was the same for both groups, any tendency toward favoring one or the other text could not be ascertained; furthermore, the test was found to have a more difficult level of readability--11-12 grade and college--than the text materials, making analysis difficult); administration (experiment occurred in regular classes with regular instructors and with "experimental" atmosphere, which might have raised motivational level to an unusual degree). A good longitudinal study (which is rare), with several additional interesting findings: attitudes of those reading easy text tended to improve over time and those reading hard text to decline over time, with attitudes of regular (non-experimental) students to remain the same over time; also, the use of an easier text did not result in "watered-down" achievement scores on the standardized test (scores were slightly but not significantly higher).

Denbow, C. J. <u>An experimental study of the effect of a repetition factor on the relationship between readability and listenability</u>. Unpublished doctoral dissertation, Ohio University, 1973.

Rewrote two 264-word passages on different topics (gun control in Great Britain, and cotton prices in the United States), from original level of 13-15 grade to 5-6 grade, as measured by the Dale-Chall readability formula. Developed an every-fifth cloze version for each passage, and administered in an information-gain paradigm: cloze version pretest of passage; delay of 6 days; reading of complete text of same passage; cloze version posttest of same passage. Assigned 140 college students (primarily freshmen and sophomores), randomly to one of four presentations for 1.5 minute presentation of complete text: normal written (<u>S</u>s read at own rate); oral (<u>S</u>s heard message read at typical rate); paced written (<u>S</u>s read individual sentences presented for same length of time as orally presented sentences); control group (read irrelevant passage, but at same grade level as experimental passage). <u>S</u>s were given 15 minutes for cloze version and also rated the interestingness of the title during pretest and of entire story during posttest. Results indicated that there was a significant difference in information gain for normal written, oral, and paced written compared to control group, but no significant differences between experimental groups. There were also positive correlations between interest ratings and information gain for both titles and stories, but the relationship was significant only in the case of the more interesting story content. In addition, the effect of readability changes upon information gain was highly significant for all experimental methods of presentation, and in addition, information gain was significantly greater for the non-preferred (cotton price) than the preferred (gun control) content. Also, there was a tendency (not amenable to statistical evaluation) for information gain to be greater for groups tested away from regular classrooms and not by regular instructors. Possible contributors to positive readability results: rewriting (very carefully done, with wide difference in versions, and with versions matched closely for style and/or content); testing (used information gain paradigm, avoiding criticism for use of only one cloze version); administration (careful control of time limits and allowance of only one exposure to message); level of motivation (related information gain to story interest-value, and presented some material in regular classrooms by regular instructors and others not). A very carefully done study.

Doyle, M. Readability as a key for evaluating junior college freshman English anthologies. Unpublished doctoral dissertation, University of Southern California, 1961.

Prepared three passages approximately 400 words long and on different topics (personal problems, science, and gardening), in three versions at approximate Flesch Reading Ease grade levels of 10, 13, and 16. Changed both word length and sentence length, but avoided other text changes. Administered to 150 junior college freshmen, allowing 10 minutes for reading a passage and answering 10 five-choice questions (without reference to the text). Found significant differences in expected direction for readability levels, with no significant differences between topics. Possible contributors to positive results: careful rewriting; rather large differences in readability levels; carefully developed test; fairly short time for reading and testing with no reference to text during testing; avoidance of special motivation. A very carefully done experiment.

Drake, L. C. The effectiveness of a selected readability formula in the prediction of student success with technical and nontechnical reading materials. Unpublished doctoral dissertation, University of Missouri, 1966.

Prepared four "information sheets" of approximately 1,000 words each by selecting (and slightly rewriting) original technical passages (on rockets and power mechanics), and nontechnical passages (on space travel and occupational information), from tenth-grade level to fifth-grade level, using suggestions of Dale and Chall and the Dale-Chall formula for measurement. No time limit given for reading, but recorded reading times; then, administered test with 25 four-choice items (on each content), to 549 seventh-grade science students. Some test score differences were significant, but most were not; no retention test score differences (after 4-week interval) were significant; reading speed differences were in expected direction, but were not significant. Possible contributors to negative results: rewriting (rather mechanical cutting of sentences, especially in "ands"); tests used (extremely large variances, in some cases larger than means); retention interval (4 weeks too long for detailed knowledge); high motivational level (tests given in regular classes by regular teachers, reading was timed, and passages were tested in sequence). However, was careful to hold human interest constant, and to use matched groups and pretest, posttest arrangement.

Feldman, M. E. The effects of learning by programmed and text format at three levels of difficulty. Unpublished doctoral dissertation, Cornell University, 1964.

Rewrote the original version of a beginning psychology program of approximately 250 words from the "some high school" to harder "some college" and "college" levels, according to the Flesch Reading Ease formula, by changing to harder synonyms and by lengthening sentences. Also compared programmed versus text format. Prepared two different cloze tests by deleting every third content word (noun, verb, or critical adjective), with deletions the same for all levels of readability (difficulty). Administered pretest (one cloze test) on one regular class day; then text on next day; then posttest (other cloze test), transfer test (analysis of two hypothetical scientific experiments), and "affect inventory" (four questions covering judgments of interest, difficulty, enjoyment, and comparisons to other texts) on third day. Subjects consisted of 72 college sophomores with high SCAT (School and College Ability Test) verbal scores and 72 with low SCAT verbal scores. Found significant differences owing to readability (level of difficulty) and to verbal ability, but not to format, on both pretest and posttest cloze scores. Found no significant differences in gain cloze scores (posttest minus pretest) for readability, verbal ability, or format. Found significant difference favoring text format over programmed format for low verbal ability Ss on transfer test (other differences not significant). On affect inventory, found that low verbal ability Ss made appropriate judgments of difficulty for the three readability levels (other differences not significant). Possible contributors to mixed effects of readability variable: rewriting (word length changed from approximately 250 words to 300 to 350 words as difficulty increased; to some extent, also, sentence length changes were a function of adding "ands"); test used (cloze test used only content words and only one of three possible forms; question of viability of gain scores arises from finding no differences owing to level of verbal ability, normally a powerful variable); conditions of administration (tested in regular classroom, without control of time, which probably raised level and effectiveness of motivation). Author indicates that other factors which may contribute to failure of gain differences owing to readability are fact that text has many supplementary materials (adding to redundancy) and that "compensation" effect found earlier by Klare and associates for reading difficult materials may be a factor (supporting the possibility suggested above).

Felix, J. L. The development and evaluation of a prescription for more readable reporting of research in guidance and personnel work. Unpublished doctoral dissertation, University of Cincinnati, 1967.

Prepared a prescription for readable writing in The Personnel and Guidance Journal consisting of maximums of: word length, 150 syllables per 100 words; sentence length, 19 words; paragraph length, 5 sentences; section length, 3 paragraphs; division length, 3 sections. Flexibility for syllable count and words per sentence was to be allowed within these maximums. A shield was developed to reveal four lines of type at a time for view. Two articles of approximately 2,400 words and 1,875 words on counselor role and counselor self-concept were "adjusted" in word and sentence length as close as possible to the above standards by members of a University of Cincinnati Evening College composition class (actual grade levels not given for readability, however), and in other ways by the experimenter. Copies of the original and rewritten versions were sent to 200 randomly selected Journal subscribers with 10 rating scales, and 44 responded. No differences between the original and rewritten versions were significant (ratings for clarity and conciseness particularly, however, and the grand mean for all ratings, favored the original version somewhat). When the above responses were received, a "content questionnaire" composed of 20 four-choice items (10 on each content) were sent to the respondents. From the returns, 25 usable questionnaires were received and analysis in one case based on 23. None of the comprehension differences were significant, but all except one favored the rewritten version. Possible contributors to negative results: rewriting (major change was in sentence length with approximately 90% of desired standard reached, whereas in vocabulary only approximately 25% of desired standard was reached; ratings by readers suggest the rewritten version was less clear than the original; rewritten version of first article was longer than original); test used (the "content questionnaire" was very short for length of articles); number of returns (N was particularly small for comprehension testing). However, this experiment was unusual and interesting as a test of readability in practice, and it seems possible that a less "conservative test" (author's words) based on better writing and testing and larger N might produce positive results, since differences were tending in that direction.

Froelich, D. M. A comparison of two methods of assessing textbook readability of selected college level electronics textbooks. Unpublished doctoral dissertation, University of Missouri, 1970.

Rewrote a 2,000-word passage on electronics from original Flesch Reading Ease level of 15th grade to 9th grade, making relatively greater changes in sentence length and relatively smaller changes in vocabulary. Developed a comprehension test of 32 four-choice items over the 2,000 words and also one every-fifth cloze version covering 500 words (except that where a deletion fell at a number or special figure, the following word was deleted instead). Administered tests to 58 college freshmen in a basic electronics course in three state colleges in Missouri, giving cloze tests on first and second days, then having Ss read entire passage and take multiple-choice test on third day. Concluded that the Flesch Reading Ease formula was not a good measure of readability because: (1) there was no significant difference in multiple-choice comprehension scores between 9th and 15th versions; (2) there were significant differences in cloze scores between the first and second half of the passage, but the Flesch Reading Ease scores failed to show this. The cloze procedure was judged a better measure of readability than the Flesch Reading Ease formula because scores corresponded more closely to scores on the multiple-choice comprehension test than did Flesch scores. Possible contributors to negative results: rewriting (change in sentences in rewritten version approximately three times as large as changes in vocabulary; also, rewriting was done in 100-word segments); testing method (used only one version of cloze test, and covered only part of the material; also, modified the standard every-fifth pattern in order to skip potentially troublesome blanks); testing procedure (tested in regular classes by regular instructors, with very liberal reading and testing times; also gave highly motivating instructions indicating that project was designed to ascertain Ss' ability to comprehend written material). However, used desirable controls for content (four judges were used to assure that technical and conceptual content were unchanged) and testing (text was removed during testing). A final factor had an uncertain effect: testing with cloze procedure just prior to reading the same passage for testing with the multiple-choice test may have raised the level of content familiarity enough so that differences in multiple-choice comprehension became relatively unlikely.

Funkhouser, G. R., & Maccoby, N. Study on communicating science information to a lay audience, Phase II. Report based upon a study funded by the National Science Foundation (NSF GZ-996). Institute for Communication Research, Stanford University, September, 1971.

This 3-year study (Phase II), was preceded by a 2-year study (Phase I), aimed at identifying the textual variables related to effective communication of science information. Phase II was designed to see if the results of Phase I could be used actually to produce effective science writing. Both Phase I and Phase II are enormously detailed. This abstract covers only those aspects of Phase II rather directly related to readability variables; the reader is encouraged to examine the complete reports of both Phases I and II for other desired details. Prepared experimental versions on three science topics, enzymology, polymer chemistry, and plasma physics. Began with preparation of 10 four-choice questions for each topic, then prepared information modules to answer these questions, and finally knitted these modules together into articles of approximately 1,000 words. The articles were then varied in terms of 20 "gross variables" (those, such as sentence length, involving entire article), and 10 sets of "information-gain variables" (each set comparing two alternate ways of presenting same information). Subjects received a "package" consisting of: an introductory sheet, a general information and background questionnaire, a "Science in General" attitude scale, and instructions for reading the included version on each of the three topics, followed by the 10-item information test. Subjects were: 475 junior college students tested in lieu of scheduled class meetings; 447 university students tested on a paid volunteer basis; and 272 scientists tested at place of employment over a period of several weeks. Testing was not timed; average time taken was approximately 45 minutes, with few taking as long as 1 hour. Results indicated that, taken singly, high and low incidence of none of the following produced significant differences in information score for the college audiences: sentence length (from approximately 15 words in easy version to 28 in hard); percentage of non-Dale words (from about 25% in easy to 35% in hard); percentage of activity words (from about 4% in easy to about .2% in hard); and science words (from about 10% in easy to 20% in hard). None affected any of the "attitude" measures either, except that percentage of science words affected judged difficulty for one group. However, combining the above variables into "easiest" and "hardest" versions produced a significant difference in information for the junior college group, as well as significant differences in judgments on attitude index, enjoyment, and difficulty estimates for both junior college and university groups. The "easiest" and "hardest" versions did not produce significant differences for the scientists on any of the above variables. The mixed results suggest that certain factors could have contributed to positive and others to negative results. Possible contributors to positive results include: careful writing (articles carefully planned, with help of scientists); testing (tests short but well designed); administration (for junior college students, out-of-class presentation may have produced a rather

(continued)

Funkhouser, G. R., & Maccoby, N. (continued)

typical motivational level). Possible contributors to negative results include: conditions of testing (for university students, using paid volunteers may have produced a rather high level of motivation; combined with liberal reading, rereading, and test time, plus chance to refer to text during testing, this may have washed out differences); level of background information (high level of background information for scientists may have precluded differences on particular test used). Despite the effort expended on this study, it is notable that the two college groups and the scientist group were tested under such different conditions, especially since motivational level could so clearly have had an effect (owing to liberal time limits and opportunity to refer to text during testing). This may have contributed to the need to combine variables that did not have a significant effect taken singly (which was noted for the information-gain as well as science information measure). Also notable was the observation that those with high self-ratings on science interest and/or knowledge were much superior on the information test to those with low.

Geyer, J. R. The cloze procedure as a predictor of comprehension in secondary social studies material. Unpublished doctoral dissertation, University of Maryland, 1970. (The study by Geyer, J. R., and Carey, A. R.--Predicting and improving comprehensibility of social studies materials: the roles of cloze procedure and readability adjustment, Reading World, 1972, 12, 85-93--appears to be identical to this dissertation. Since the dissertation provides more detail than the paper, this abstract was based upon the dissertation.)

Selected sections of approximately 1,500 words on the same topic (the depression of the 1930s) from two junior high school American history books, one section rated at 5-6 grade and the other at 7-8 grade by the Dale-Chall formula. Developed a cloze version over 250 words of each section, deleting every fifth word beginning with word 1. Also developed a multiple-choice test of 40 four-choice items. Administered to 201 eighth grade students in the following order: take cloze version of text, read complete version of text, take multiple-choice test. Concluded that cloze test did not predict student comprehension of social studies material better than (1) standardized reading test scores, or (2) I. Q. scores, or (3) previous social studies grades, but that results warranted further investigation. Also concluded that the more readable text did not result in higher comprehension than the less readable. Possible contributors to negative readability results: text used (experimenter did not rewrite text, but instead used text on same topic from two books; gave only sample from each of the two books, and did not describe how the two differed in readability variables); tests used (used same multiple-choice test to cover both books; used only one version of cloze test, and not the one generally felt most desirable); differences in readability (compared versions differing only between 5-6 and 7-8 in grade level, and based on different books); relatively high and effective level of motivation (tested regular classes, with very liberal time limits for reading and testing; scores on multiple-choice text complicated by previous reading of cloze version shortly before).

Gilman, D. A., & Moreau, N. A. Effects of reducing verbal content in computer-assisted instruction. AV Communication Review, 1969, 17, 291-298.

Rewrote a computer-assisted instruction program on topic of significant figures, changing: (1) long, complex sentences to short, concise ones; (2) unusual words to common words; and (3) length of verbal content of feedback messages as much as possible without losing meaning or content. No readability scores are given before or after. Developed a pretest of 5 items to test prior knowledge of topic; also developed a posttest of 18 multiple-choice items, including items to measure mastery and transfer. Administered each version to 18 community college students (naive with respect to educational experimentation) via IBM 1050 terminals, with time needed to complete program recorded by computer. Results showed: (1) no difference in posttest scores; (2) significant difference in time needed to complete program favoring easier version; and (3) no difference in magnitude of correlation between Ss' California Test of Mental Maturity verbal intelligence scores and posttest scores for harder versus easier version (though value was higher for harder version than for easier). Possible contributors to mixed results: rewriting (changes not specified exactly, but included changes in length); testing (conditions not specified exactly, but participation in experiment may have affected motivational level of naive subjects); administration (greater time needed for completion of harder version may have been due to slower type-out time because of greater length).

Hayes, G. W. The relationship of socio-economic status of pupils to their comprehension of reference materials written at different levels of readability. Unpublished doctoral dissertation, University of Illinois, 1966.

Rewrote a passage of approximately 1,000 words entitled, "Caring for the Sow and Litter at Farrowing Time," from a Flesch Reading Ease Score of 66 (approximately 8.5 grade level), to one of 91 (approximately 5.8 grade level). Developed a 30-item four-choice test, with 10 items on understanding, 10 on recall of facts, and 10 on application. Administered test to 96 ninth-grade pupils enrolled in vocational education in agriculture courses in 21 Illinois high schools, with testing done in regular classes. Ss were classified into socio-economic classes. Found significant differences favoring the more readable version on the recall portion of the test and favoring the middle-working-class group on the comprehension portion of the test. There were no significant differences on the application portion or on total test scores. Possible contributors to mixed results: difficulty of materials used (original version was already at approximate reading level of Ss, with rewritten version lower); level of motivation of Ss (testing took place in class, and was untimed). However, some careful controls were used (original author okayed content of rewritten version; only one reading was permitted, with text removed during testing).

Hiller, J. H. Learning from prose text: Effects of readability level, inserted question difficulty, and individual differences. Journal of Educational Psychology, 1974, 66, 202-211.

In this study readability was only one of the variables used in a rather complex design. The abstract below concentrates on the effects of readability upon student behavior; a reader wishing detailed information on the other variables covered should consult the study itself. Rewrote a 1,687-word section of a college-level text covering modern mathematics. Prepared two versions by breaking long sentences into shorter ones and by replacing uncommon terms with more familiar synonyms. Did not measure readability levels, but three versions were: original, 1,687 words--low readability, 22.0 words per clause unit; moderate readability, 1,874 words--18.4 words per clause; average readability, 1961 words--14.9 words per clause. In latter two, also added a concrete example to clarify an abstract discussion. Developed 28-item four-choice test; allowed 25 minutes to study text, then removed text and asked for judgments, and then allowed 20 minutes for testing. Also gave test on open-book basis to some students, and as a two-week delayed retention test to some students. Results indicated that readability versions produced: (1) significant differences in comprehension in expected direction using open-book test and using closed-book test (immediate retention); (2) significant differences in subjective ratings, in expected direction; (3) significantly higher interest ratings in expected direction; (4) nonsignificant differences in delayed retention testing. (Other readability effects were confounded with other treatments, and will not be reported here.) Possible contributors to positive results: rewriting (appears to be large differences in readability versions); administration conditions (Ss not told test would be open-book); and test used (careful construction and long test). Possible contributors to mixed results: administration conditions (text present during open-book testing in some cases; however, test was not present in other cases, and study time was rarely fully used in either situation). An interesting study, especially since ratings for readability and interest also followed tested differences between readability versions. Also of interest is difference in immediate retention scores between very difficult and moderately difficult but not average difficulty lessons. Lack of readability scores on lessons and use of other variables in addition to readability to make interpretation uncertain, however.

Hites, R. W. The relation of readability and format to retention in communication. Unpublished doctoral dissertation, The Ohio State University, 1950.

Wrote four paragraphs on map reading at three levels of difficulty, and also varied them in paragraphed versus unparagraphed form and with and without paragraph headings. The readability variations were produced by changing proportion of words not on the Dale 3,000-word list and by changing sentence length, resulting in a 918-word passage at approximately 7-8 grade level, a 916-word passage at approximately 11-12 grade level, and an 857-word passage at approximately college (13-15) grade level, according to the Flesch and Dale-Chall readability formulas. Developed a comprehension (or retention) test of 20 four-choice or five-choice items written at approximately 11-12 grade level. Administered test to 771 college freshmen in ROTC classes, with regular instructors conducting experiment during regular class hours. Allowed 5 minutes for reading, then removed text during testing; no time limit given for testing. Obtained Ohio State Psychological Examination scores for all Ss. Analyses of variance showed significant variance attributable to intelligence and format (unparagraphed versus paragraphed), but not to level of difficulty (readability). Found correlations between Ohio State Psychological Examination scores and comprehension scores significantly different from zero for all nine classes, but correlations at the most difficult text level were not consistently higher than those at the least difficult level, as had been hypothesized. Possible contributors to negative results with readability variables: levels of rewritten material (the most difficult level was only slightly higher than the reading level of Ss); conditions of administration (motivation level of Ss apparently raised by having regular instructors conduct experiment in regular classes, with motivation allowed to have an effect through liberal reading time and uncontrolled testing time; author indicated, furthermore, that Ss "would be motivated to do their best on the experiment since map reading was part of their course"). Author noted that, since differences between levels were in the expected direction, use of Ss with lower intelligence or mental age may have produced expected differences (it seems possible that motivational effects may have been equally or more nearly responsible). Carefully done study but note that length of readability versions varied slightly, and judges were not used to determine content equivalence of versions.

Johnson, K. H., Relova, R. P., Jr., & Stafford, J. P. An analysis of the relationship between readability of Air Force procedural manuals and discrepancies involving non-compliance with the procedures. Master's thesis, Air Force Institute of Technology, Air University, August, 1972. Document No. AD 750 917. Reproduced by National Technical Information Service.

Reviewed literature which showed that military documents are not written at a readability level commensurate with intended readers' ability to read them. Investigated question of relationship between the amount of readability/reading ability gap and the frequency of non-compliance discrepancies with procedures. Examined data in the Air Defense Command discrepancy file for years 1970 and 1971, yielding 25 usable cases for 1970 and 14 for 1971. Used Flesch Reading Ease formula to determine readability levels of documents. Examined reading ability of target population, using data from the Uniform Military Record for each Air Force Specialty involved. Selected 90% cumulative grade level (reading level exceeded by 90% of the population) as a measure because it is a function of both mean value and variance for each population. Ran Spearman Rank Correlation measures on all categories of data separately for years 1970 and 1971 and for both combined. Results indicated: (1) no direct relationship between readability of documents and frequency of discrepancies; (2) consistent and significant relationships between target population's reading ability and frequency of discrepancies; and (3) a direct and significant relationship between the readability/reading ability gap and frequency of discrepancies for the 1970-71 data combined, with separate data for 1970 and 1971 in the same direction (possibly not significant owing to small *Ns* involved). Authors conclude that the gap is directly related to non-compliance with procedures, and estimate that approximately a 40% reduction in errors could be achieved by eliminating the gap. Interesting relationship found, especially considering the necessarily crude nature of the data involved; however, the field-study nature of this test may also have contributed to positive results owing to "typical" level of motivation involved.

Kincaid, J. P., Yasutake, J. Y., & Geiselhart, R. Use of the automated readability index to assess comprehensibility of Air Force Technical Orders. Technical Report SEG-TR-67-47. Wright-Patterson Air Force Base, Ohio: Systems Engineering Group, November 1967. (Appears almost identical to: Smith, E. A., and Kincaid, J. P., Derivation and validation of the Automated Readability Index for use with technical materials, Human Factors, 1970, 12(5), 457-464. Since the only differences are in the descriptive information on the Automated Readability Index, and not in the experimental portion of the paper, both may be considered abstracted below as far as the latter portion is concerned.)

First part of paper describes the Automated Readability Index (ARI), consisting of an attachment for an electric typewriter that tabulates number of strokes, words, and sentences in material being typed and derives from these ARI indices that either may or may not be in terms of graded difficulty levels of the material. Second part describes an experiment involving rewriting of two technical passages of approximately 250 words each on the electrical and power systems of C-141A aircraft. Rewrote passage at 3 levels, approximately 16, 12, and 8 grades according to ARI formula, using traditional readability variables; judges determined that technical meaning, content, and terms were not changed. Developed test of 8 questions (type not specified, but probably multiple-choice) for each passage. Administered to 120 male airmen in mechanics courses, all of whom were recent high school graduates. Found significant effects for readability, passages, and interaction; within readability versions, found easy (8 grade) and medium (12 grade) versions significantly better comprehended than hard (16 grade) version, with no difference in raw scores between easy and medium. Possible contributors to positive aspect of results: rewriting (carefully done, with large gaps between grade levels of versions); testing (used pretesting to develop clear testing procedure and to item-analyze test). Possible contributors to negative aspect of results: administration may have tended to yield high level of motivation (tested in regular classroom, and gave Ss time and opportunity to refer to text during testing). Carefully done study, with use of technical experts as judges of content and technical term equivalence; positive effect of large gaps in readability version apparently partially counteracted by negative conditions of administration. Third part of paper describes characteristics of the ARI (reliability, validity, ease of application, and limitations).

Kincaid, J. P., Van Deusen, J., Thomas, G., Lewis, R., Anderson, P. T., & Moody, L. Use of the Automated Readability Index for evaluating peer-prepared materials for use in adult reading education. Final Report, Project No. 1-D-054, Grant No. OEG-4-71-0069. Statesboro, Georgia: Georgia Southern College, September, 1972.

First part of paper describes peer-prepared stories, collection of stories, and rating by reading specialists and trainees. Second part, the only one described in detail below, describes experiment using rewritten versions of three best stories. Third part compares Automated Readability Index (ARI) with Flesch Reading Ease and Gunning Fog Count as to test-retest reliability and application time. In second part of study, rewrote three best peer-prepared stories (on topics of fog, bike ride, and foot race) at grade levels of 4, 8, and 12 according to ARI, using traditional word and sentence alterations (with length varied slightly). Developed an every-fifth cloze version of each passage (apparently only one) and administered to all students in remedial reading courses in upper high school grade level (apparently done individually, since some Ss did not complete work); reading ability of Ss on Gates-MacGinitie reading test ranged from 3.1 to 10.4, with mean of 6.23. Accepted synonyms as correct in addition to misspellings. Found significant effects for readability, passage, and interaction on percentage correct scores. Further analysis showed significant differences between easy-hard and medium-hard versions, but not between easy-medium; also foot race passage found significantly harder than other two passages. Considering mixed results, possible contributor to positive aspects include: rewriting (large gaps in grade level between versions); and, administration (apparently tested on ad libitum basis, without motivational pressure). Possible contributors to negative aspects: tests used (apparently used only one cloze form); passages selected (use of best passages may have reduced effect); scoring used (acceptance of synonyms may have introduced unreliability). Finding of high and significant correlations between Ss' cloze scores and reading test scores is valuable.

Kincaid, J. P., & Delionbach, L. J. Validation of the Automated Readability Index: A follow up. Human Factors, 1973, 15(1), 17-20. (This study is an extension of the earlier 1961 Kincaid-Yasutake-Geiselhart study--and the Smith-Kincaid study of 1970--with a different, heterogeneous group of subjects but the same materials.)

Rewrote (earlier) two technical passages of approximately 250 words each on the electrical and power systems of C-141A aircraft at three grade levels, 16, 12, and 8, according to Automated Readability Index (ARI) formula. Used traditional readability variables, with judges determining that technical meaning, content, and terms were not changed (in earlier work). Used eight test questions developed (earlier) for each passage. Administered to 110 male enlisted men in National Guard; age, service, and education varied widely (in case of latter, from 7th grade education to doctorate). Found significant effects for readability and passage variables; within readability versions, found easy (grade 8) and medium (grade 12) versions significantly better comprehended than hard (grade 16) version. Possible contributors to positive aspect of results: rewriting (carefully done, with large gaps between grade levels of versions); testing (had earlier done careful pretesting to develop sensitive test and procedure). Possible contributors to negative aspect of results: administration (Ss given ample time for reading and testing and allowed to refer to text during testing--but not turn back to other passage). Results very similar to earlier study (see above). An additional interesting result was finding of significant correlation between Armed Forces Qualification Test scores of individuals and their comprehension scores on the hard versions but not on easy or medium versions.

AD-A042 335 NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
SURVEY OF TECHNICAL MANUAL READABILITY AND COMPREHENSIBILITY, (U)
JUN 77 T E CURRAN
UNCLASSIFIED NPRDC-TR-77-37 NL

2 OF 2
AD
A042335

END
DATE
FILMED
8-77
DDC

Klare, G. R., Mabry, J. E., & Gustafson, L. M. The relationship of style difficulty to immediate retention and to acceptability of technical material. The Journal of Educational Psychology, 1955, 46, 287-295. (This study was one of a series reported in greater detail in the report, "The relationship of verbal communication variables to immediate and delayed retention and to acceptability of technical training materials." San Antonio, Texas: Air Force Personnel and Training Research Center, Lackland Air Force Base, December 1954--Research Bulletin, 54-103. This abstract is based upon the published article, with some relevant detail added from the research report, however.)

Rewrote a 1,206-word lesson with the first half on the induction system and the second half on the cooling system of a reciprocating aircraft engine as used in an aircraft mechanics training course. Original (present) version, at style difficulty level of 11-12 grade according to Dale-Chall and Flesch Reading Ease readability formulas, was rewritten to Easy and Hard style difficulty levels, at 7-8 grade and 16+ grade, respectively, by changing typical readability variables (percentage of short, familiar, frequently used words; average sentence length, and the related proportions of simple sentences and prepositional phrases; the proportion of abstract to concrete wordings; and, the proportion of active to passive constructions). Technical experts served as judges to determine that neither content nor technical terms were changed. Prepared "unsplit" versions with both halves of the lesson having the same style difficulty level and "split" versions with the halves having two different levels. Numbered every fifth line of all versions for later measurement of amount read by subjects. Developed a 50-item, multiple-choice test over the standard (present) version for use with all three levels of style difficulty. Administered test to 989 male airmen in indoctrination training, allowing 20 minutes for reading and rereading. Ss indicated amount read (line being read when time was up), judged which half of this versionwas "easier to read" and "more pleasant to read," and were then given 40 minutes for testing (text not present). Results indicated that readability had: (1) a significant effect upon number of words read (Easy and Present levels significantly more than Hard), but not number of lines read; (2) a significant effect upon judgments of easier and more pleasant to read (Easy and Present levels significantly more than Hard) with content equated (using split versions); and (3) a significant effect upon immediate retention (comprehension) test scores (Easy mean score higher than Present or Hard, with Present higher than Hard but not quite significantly). Further results showed that: (1) content differences (preference for cooling system over induction system content) over-rode style differences when the two were confounded; (2) Ss having the Easy style level answered significantly more items correctly than those having Present who answered significantly more than those having Hard; and (3) delayed retention test scores (after two weeks) showed no significant differences except that those having had a harder style level in the first half of a split version retained significantly more test information from the second half of the version than those having an easier first half (this effect was also shown

(continued)

Klare, G. R., Mabry, J. E., & Gustafson, L. M. (continued)

in the immediate retention test scores, but did not reach significance). Preliminary work, using the same materials but involving airmen in advanced technical training classes, showed no comprehension differences, which was attributed to the demonstrated high level of test knowledge before reading text (test scores reached over two-thirds correct before reading). Possible contributors to positive results: rewriting (careful revision, including use of judges to determine equivalence of content and technical terms); tests used (determined that questions were spread evenly over content; seventh and final form of test had high reliability; measured amount read and preferences; also, use of split versions made possible equating for content); administration (large numbers of subjects, with text removed during testing). Further analysis now suggests that negative results during testing in advanced classes may have been influenced by high effective level of motivation (testing in classes with liberal reading and testing time) as well as level of background knowledge of Ss. Experimental design aided interpretation by removing any possible effects of ability differences in groups or bias of test toward easier or harder styles.

Klare, G. R., Shuford, E. H., & Nichols, W. H. Relationship of style difficulty, practice, and ability to efficiency of reading and to retention. Journal of Applied Psychology, 1957, 41, 222-226.

Used two variations of a 75-76 word passage prepared from a standard lesson used in the training of Air Force reciprocating engine mechanics. The Easy style difficulty version, at approximately 6-7 grade level and the Hard version, at approximately 15-16 grade level (as judged by the Dale-Chall and Flesch Reading Ease readability formulas) were developed by varying: (1) the percentage of short, familiar, frequently used words; (2) average sentence length (in words); and (3) the proportion of concrete to abstract wordings. Technical experts determined that neither content nor technical terms were changed from the standard version. Two retention tests were used: (1) a modified recall test consisting of four sentences with two blanks each (sentences not drawn directly from standard text, but words for blanks common to Easy and Hard style difficulty versions); and (2) a recognition test of 18 words (9 common to the two versions and 9 from other technical sources, graded on a right-wrong basis). Administered passages, using an eye-movement camera, to 120 male airmen in a mechanics course, half high and half low in mechanical aptitude, with reading time and visual fixations recorded. Developed a "strong set to learn" by using a difficult preliminary passage prior to the experimental passage, then allowed one or three readings of the experimental passage. Results indicated that all mean values were in the predicted direction, i.e., style difficulty (readability), practice and ability: (1) produced differences in the words read per fixation measure and the words read per second measure (the effect of readability upon the latter, however, did not quite reach significance); and (2) produced differences in the recall and recognition measures (the effect of readability upon recognition, however, did not quite reach significance). Possible contributors to positive results: (1) rewriting (careful revision, with technical experts used as judges); (2) tests used (recall test was expected to produce more clear-cut results than a recognition test; also, inclusion of time and fixation measures provided a seldom-used basis for interpretation of results); (3) procedures used (eye-movement camera provided precise control of reading, as did development of a strong set to learn). Of additional interest in this study were the findings that: (1) the effect of readability did not appear in the words per fixation measure under a "weak set to learn," but only a strong set; (2) more able Ss profited more from added reading (practice) than less able; and (3) the effect of readability upon reading efficiency and learning was as clear-cut as upon comprehension measured in a more traditional experimental setting.

Kulp, M. J. The effects of position practice readability level on performance. Unpublished Masters thesis, San Jose State University, 1974.

Editors prepared three versions of a "position practice" (a document containing step-by-step procedural instructions to be followed in learning and performing a job) at grade levels 5, 10, and 15 as measured by the Gunning Fog Index. Edited training booklet to accompany the position practice at same three levels; both position practice and training material were preceded by a document telling how to read a position practice written at level 5. Administered to 72 Ss in three groups of 24 matched on Gates-MacGinitie Reading Test scores, with all paid and a bonus promised to approximately the top half. Ss first given uncontrolled time to read the document telling how to read a position practice and take the fill-in test covering it (mean number of errors was very small--5%). Next, Ss were given uncontrolled time to read training material (training booklet and position practice), but time actually needed was noted. Finally, performance test consisted of a set of 40 "Status Change Notices" to be processed into "Payroll Change Notices," and a new position practice for this purpose (without training material), being also asked to record time needed for separate "Status Change Units" to be processed (only 9 Ss completed performance test, with mean number of 21.2 Payroll Change Notices completed). Analyses of variance of data indicated that for all three categories of dependent variables, training time scores, performance error scores, and performance time scores, the effect due to reading level (RL) of Ss was significant but the effect due to document readability level (DRL) and the interaction of the two were not significant. Regression analyses made for all three categories involved the difference between DRL and RL (i.e., the gap or mismatch) for each grade level of material. These data indicated, in all but two inconsistent areas, strong support for the hypotheses that training time, performance errors, and performance time will increase as the document reading level exceeds the S's reading level, with the relationship apparently linear in all but one case. The two exceptions were for performance time analysis for 10th grade material, where a higher degree regression analysis was needed, and training time for 15th grade material. In the latter case, it was found that four Ss with low reading ability apparently began "flipping" pages in order either to keep up with other Ss or because they "gave up" when faced with the difficult reading task; this hypothesis received support from sequential training time scores for units, where the time taken changed from a greater value to a lesser value than that for other 20 Ss. Furthermore, the mean error score for the four was 11.7 per Payroll Change Notices, comparedto 2.8 for the other 20 Ss. Other analyses showed that where the reading level scores of Ss were within two grade levels of the document reading level (i.e., the gap or mismatch), an average of two errors was made per Payroll Change Notice), while for a gap of two or more grade levels the average was 10 errors; the author estimates that $33,288 could be saved in error correction costs on this basis (approximately 79%). Corresponding

(continued)

Kulp, M. J. (continued)

performance time means for the small and large gap were 8.15 minutes versus 12.24 minutes for each Notice (a saving of approximately 33%). Corresponding overall training time means for the small and large gap were 3.04 hours versus 3.42 hours (a saving of approximately 11%). The author explains the smaller saving in terms of some Ss giving up and not benefitting from the training material when their reading ability was far below the readability level of the document, and urges performance error measures (where the saving was larger) as the better index for this and similar studies. A complex but carefully done study. Possible contributors to positive results: materials used (careful selection and preparation of experimental materials); subjects (careful matching of Ss in groups by reading ability); measures used (error as well as time measures used); length of study (a day-long experiment, where effects of high level of motivation may tend to wear off); statistical analyses (use of regression analyses). However, several items may have affected results: rewriting (not specifically stated what editors actually did in rewriting except achieve three formula grade levels); very high level of effective motivation (Ss paid and given bonus for good work, which, plus uncontrolled reading and performance time, may help to account for lack of differences in analyses of variance; however, this may have been felt necessary owing to length of experiment, and may have been felt desirable as a good field-test under fairly realistic conditions; also, use of easy grade 5 passage and easy test before experimental situation may have developed a weaker set to learn which may have counteracted somewhat the effects of the high level of motivation, as the "group pressure" to finish early may also have done, at least for the time measures).

Laubach, R. S. A study of communication to adults of limited reading ability by specially written materials. Unpublished doctoral dissertation, Syracuse University, 1963.

Rewrote four articles (total of approximately 1,000 words) taken from Christian Science Monitor from original ninth-grade level to approximately 2.7 grade level, using readability theory (one version), plus added linguistic controls (two other versions). Administered a comprehension test of 10 multiple-choice items for each article (mean reading grade level of tests was 2.7), to 164 adults in Philadelphia and 104 in Cleveland adult elementary reading classes. Found significant difference favoring rewritten versions over original; adding linguistic controls to readability version did not further increase comprehension, however. Possible contributors to positive results: large, carfully written difference in readability (6+ grades); adequate test; relatively short reading and testing time; motivation not specially raised by experimental conditions. Tested in class, but during four periods and following warm-up period.

Major, A. G. Readability of college general biology textbooks and the probable effect of readability elements on comprehension. Unpublished doctoral dissertation, Syracuse University, 1955.

This study had three major parts, described sequentially below. Part I. Prepared three biology passages on epithelial tissue: No. 1 (445 words), was at Flesch Reading Ease (RE) score of 36 (approximately 14th grade level), 2 (692 words) at RE of 46 (approximately 12th grade), and 3 (713 words), at RE of 58 (approximately 10th grade). Change from 1 to 2 described as reduction of vocabulary load (sentence length also was markedly reduced, however); change from 2 to 3 described as reduction of concept load (vocabulary index was markedly reduced, in addition to number of concepts). Administered 15-item test (five 5-choice and 10 matching items), to 170 college freshmen non-majors, allowing 15 minutes for reading, but no reference to text during testing. Found significant differences between 1 & 2 and 1 & 3 for above average and average, but not below-average readers. Part II. Made nation-wide survey of general biology textbooks. Part III. Selected excerpts from three books on same topic (flowers), of almost same length (475, 374, and 431 words), at three RE levels, 44 (approximately 13th grade), 49 (approximately 12th), and 60 (approximately 10th). Difference between first and second was clearly most marked in sentence length, and between second and third in vocabulary (syllable length). Administered 20-item test (10 four-choice, 5 modified true-false, and 5 matching items), to same 170 Ss as above under essentially same conditions. Found significant differences between first and second and first and third, but not second and third for above average and average, but not below-average students. Possible contributors to positive results: very careful rewriting; good tests; though reading time was long, did not allow perusal of text during testing; motivation not specially raised. Somewhat surprising, however, to find that below-average readers (defined as those getting lower third of class grades), not helped; also, that passages on same topic, but in different books showed significant comprehension differences as predicted despite probable content differences of some sort. Very carefully done, and excellent study; note, for example, that scientific vocabulary was held constant when readability changes were made.

Marshall, J. S. The relationship between readability and comprehension of high school physics textbooks. Unpublished doctoral dissertation, Syracuse University, 1956.

Rewrote a passage on basic electricity selected from a high school physics text (first half = approximately 500 words on electrical potential and voltage; second half = approximately 1,200 words on current and resistance), from Flesch Reading Ease score of 47 (approximately 13th grade level), to a score of 65 (approximately 9th grade level). Indicated intent to change both sentence length and number of syllables per 100 words (however, most change was in sentence length measure, with only small amount of change in syllable length; also, Flesch Human Interest score was changed in rewriting). Allowed total of 14 minutes for reading (8 for first half, 6 for second), then removed text during administration of 31-item four-choice test (13 questions on first half, 16 for second). Differences in test scores between versions were in the expected direction but were not significant; readers judged the rewritten version easier. Possible contributors to negative results: rewriting changes (rewrote "to formula"; changed sentence index more than word index; shortened sentences rather mechanically, primarily by removal of "ands" and semicolons; changed human interest); testing conditions (probably raised motivation level by doing testing himself in regular class and doing reading and testing of halves in sequence; probably allowed motivation to have effect by giving fairly generous reading and testing times). However, was careful to have judges ascertain that content did not change between versions, and did not allow reference to text during testing.

McCracken, R. A. An experiment with contrived readability in fifth and sixth grades. The Journal of Educational Research, 1959, 52, 277-278.

(This study appears to be a report of the dissertation done by the author. Since the dissertation itself is not available through University Microfilms, this abstract is based upon the article with additional notes from the dissertation.) A 475-word fairy tale ("Jack and the Talking Tree"), which would normally be at third or fourth grade level, was "written up" to 7-8 grade according to Dale-Chall and 8.7 grade according to Yoakam formulas by using complex sentence structure and somewhat unnatural, difficult vocabulary. A 286-word science selection ("Polar Pacific AirMasses"), which would normally be at 7-8 grade, was "written down" to grade level 4 according to Dale-Chall and 5.2 according to Yoakam formulas by using simple sentence construction and easy vocabulary. Developed 10 factual content four-choice questions on each selection. Administered both selections to 124 fifth and sixth grade children in regular classes by regular teachers, with no time limits for reading, rereading, and testing, and with text available during testing. Results showed higher mean score (6.5) for "Jack and the Talking Tree" (less readable) and lower mean score (5.2) for "Polar Pacific Air Masses" (more readable). Ss almost unanimously judged "Polar Pacific Air Masses" to be easier to read. Possible contributors to negative results: rewritten material (comparison was between two different contents as well as style difficulties, confounding the results); tests used (no attempt was made to check on test-item knowledge prior to reading); administrative procedure (level of motivation likely to be high owing to testing in class by regular classroom teachers, and to have a marked effect owing to unlimited reading-testing time and text present during testing). An interesting result was the almost unanimous agreement that the more readable selection (according to formula) was almost unanimously judged easier to read; this suggests further that the conditions of the experiment did not provide a proper test of the effect of readability--even as far as the questionable procedure of writing material to conform to formulas (which the author calls "contrived readability") is concerned.

McLaughlin, G. H. Comparing styles of presenting technical information. *Ergonomics*, 1966, *9*, 257-259.

Rewrote a difficult 2,200-word pamphlet on damp-proofing of solid floors. Primary change was in format, but number of words also cut by approximately half in revision. Original version had mean sentence length of 28 words; revision had mean of 23 words. Prepared a 10-item test requiring mostly one-word answers. Administered original to 64 highly motivated freshmen embarking on university courses in technology; warned *S*s that they might have to take extra work in efficient reading if they performed badly. Later administered to eight randomly selected subjects who were not highly motivated; *S*s were told to work at any rate they liked since scores would not be used for individual assessment. Also administered easier revisions to 64 randomly selected freshmen who were again highly motivated, and later to eight randomly selected freshmen who were not highly motivated. Results showed that *S*s who *were not* highly motivated scored significantly higher on easier than on harder versions; difference for *S*s who *were* highly motivated was not significant. Highly motivated *S*s scored higher on both easier and harder versions than *S*s with lower motivation. This study is interesting despite lack of specific readability information in that comprehension differences between easier and harder versions appeared only under conditions of lower rather than higher motivation. (Highly motivated *S*s recognized difference between easier and harder versions; said they would not have read harder except under duress, because of its difficulty.)

McTaggart, A. C. *An experimental validation of the Flesch and Dale-Chall readability formulas on high school health texts*. Unpublished doctoral dissertation, University of Illinois, 1962.

Rewrote approximately 2,250 words of ninth-grade health material at seventh-grade and twelfth-grade levels according to suggestions of Flesch and approximately 2,250 according to Dale and Chall's suggestions. Administered 34-item multiple-choice test covering original and rewritten versions to nine classes of ninth-grade students (N = 208-210). Found scores on seventh-grade level passages significantly higher than those on ninth-grade level; scores on ninth-grade level also higher than those on twelfth-grade level, but not significantly. Found rewriting using Flesch's suggestions more effective for students high in health knowledge, whereas rewriting using Dale and Chall's suggestions was more effective for those low in health knowledge. Possible contributors to positive results: careful rewriting; adequate test; relatively short reading and testing time; motivation not specially raised by experimental conditions. However, differences between levels may have been somewhat reduced by allowing students to refer to text during testing.

McWhorter, K. T. The influence of passage organizational structure upon two estimates of readability. Unpublished doctoral dissertation, State University of New York at Buffalo, 1974.

Selected four passages of 250-300 words that had the following four different organizational structures (and topics): (1) time-order (glass making); (2) cause-effect (common cold); (3) comparison-contrast (social structure in the United Kingdom); and (4) statement-support (atheism). First applied Dale-Chall readability formula, then adjusted all passages to same score (8.5) by adding, deleting, or substituting words, or altering sentence length. Next, applied Bormuth machine computation passage level readability formula, again adjusting passages to same score value (.294). The Dale-Chall formula was then reapplied. Developed cloze versions on passages, using an every-fifth deletion ratio and preparing all five possible versions of each passage. Administered during regular English class periods, by regular instructors, to 120 freshmen at Niagara Community College; Ss were told they could have as much time as needed, and that tests would not affect course grades. Found following scores for passages (1) through (4) above: 14.37, 17.67, 19.70, 21.98. Four of the six passage combinations yielded significant differences, with time-order passage significantly more difficult than the other three. Also found that: Bormuth formula gave significantly higher estimate of difficulty than actual scores for combined passages and passages taken separately; and Dale-Chall estimate significantly higher than Bormuth estimate. Possible contributors to mixed results: passages used (since topics differed, interest or preference differences may have affected results); adjustments to passages (changes made to equate passages may have affected results); administration (high level of motivation related to testing in regular class by regular instructor, with liberal time limits).

Moir, L. H. A linguistic analysis of certain stylistic elements of selected works of literature for children and their relationship to readability. Unpublished doctoral dissertation, Wayne State University, 1969.

Prepared cloze tests (every fifth deletion ratio) on 10 passages (215 to 285 words), from five juvenile biographies about Abraham Lincoln. Administered to 28 Ss, identified as most able fifth and sixth grade readers in a suburban school, over period of 1 week. Related cloze scores to following variables: (1) T-unit length (unit developed by Kellogg Hunt--includes one main clause with all subordinate clauses attached to it); (2) Sentence Content Index (number of kernel sentences in deep structure divided by number of T-units); (3) Word Content Ratio (number of content words divided by "other"--function--words); and (4) L-unit (length of language strings referring to plot, setting, characterization, and dialogue). Found that as difficulty of written passage increases (readability becomes lower), length of T-units, Sentence Content Indices, and Word Content Ratios increase; effect of L-units not as clear-cut. Says (p. 113), "The shortcomings of earlier readability formulae is that they seem to suggest that by controlling just a few of the linguistic parts, such as sentence length, number of syllables, or occurrence of 'hard words,' the written passage can be made easier. This concept of how language operates or how it can be manipulated is too simplistic in light of present linguistic knowledge." Note the following in this study: small number of cases; only one version of cloze used (out of five possible); no attempt to relate variables used to traditional readability variables (despite assertions of the greater soundness of the former).

Moore, A. J. The preparation and evaluation of unit text materials for low-ability junior high school students. Unpublished doctoral dissertation, State University of Iowa, 1961.

Rewrote approximately 242 pages of science text on the topics of electricity, machines, and heat from 9-10 grade level to 5-6 grade level, as measured by the Dale-Chall formula. In addition to changing the typical readability variables of vocabulary and sentence length, also changed: number and abstractness of concepts presented; method of organization and presentation; directness of approach; style of composition; number of science words used (only "essential" words included); presentation and repetition of science and non-science words likely to be unfamiliar. Selected excerpts of approximately 1,000-2,000 words from the original and rewritten text as a basis for constructing a science test of 65 four-choice items. Administered text and test on two consecutive days in regular classrooms to 970 low-ability ninth-grade students. Twenty-one teachers who used the rewritten materials in classes rated them in comparison to regular text materials. Found very highly significant differences in test scores favoring the rewritten text, and teachers also rated the rewritten text as much superior. This study is unusual compared to other experimental studies of rewriting in the size of the positive differences found; this is especially surprising because the conditions of administration (testing in regular classrooms by regular instructors; liberal time limits for reading and testing, with encouragement to re-read; text available for reference during testing; rather motivating instructions) typically tend to reduce differences between original and rewritten versions. Analysis suggests that the highly positive results may have been influenced by the large-scale changes, during rewriting, of variables other than typical readability (style) variables. It seems likely that content was altered along with style (no check was made by judges for content commonality, and teachers actually noted during rating that desirable science principles and concepts were included in the rewriting and that the scope was effective for low-ability Ss); this may well have biased test results in a positive direction.

Murphy, W. F. <u>The application of readability principles to writing of task statements: Effects on reliability of job incumbent responses.</u> Unpublished doctoral dissertation, Purdue University, 1966.

Rewrote 50 statements in each of two forms of a 272-statement job inventory according to accepted readability principles, using an alternate 50 of the unchanged statements as controls and the remaining 172 to reduce memory cues (original and revised readability levels were not given). Mailed inventories to 250 Air Force Officers, then sent a second mailing of same forms 30-35 days later. A total of 98 usable inventories was returned from the two administrations (47 for one form, 51 for the other). Found differences in reliability at .01 level using Wilcoxon test (differences in Phi and Winer coefficients favored rewritten statements also, but fell slightly short of significance at .05 level). Possible contributors to positive results: careful preparation of statements (e.g., in rewriting sentences to avoid mere removal of "ands"); use of judges to be sure content was unchanged; field setting (as opposed to experimental setting) provided "typical" motivational level, and no special pressure was applied. Failure to present original and revised readability levels complicates interpretation somewhat; however, elaboration of the following rules indicates how readability principles can aid in the writing of task statements: (1) reduce total word value per task statement; (2) reduce average word value per task statement; (3) reduce the number of syllables per word; (4) reduce the number of syllables per task statement; (5) use double conjunction "and/or" to replace the conjunction "and" when a task statement is composed of parts which may be performed independently; (6) do not use technical terminology where the range of incumbents' experience may vary greatly; and (7) have all rewritten statements reviewed by job experts for proper interpretation and scoring. In addition, the following subjective rules were followed: (1) wording of a task should not be so specific as to include the recipient of work performed; and (2) task statements which appear to be very general should be rewritten in a manner which confines the task to the total job context.

Norton, W. H. Computer analysis of readability factors. Unpublished doctoral dissertation, The University of Iowa, 1973.

Described a computer program to analyze certain readability factors in educational text. Uses the American Heritage Word Frequency Book (Carroll, Davies, and Richman) as the basis for vocabulary analysis. This list of words is based upon a computer analysis of five million words from over one thousand books for grades 3 through 9 as selected by educators. American Heritage has proposed a Standard Frequency Index (SFI) as a readability index which presumably reflects both vocabulary and sentence structure factors. SFI = 10 (logc + 10), where c represents the fraction found by dividing the counted word frequency into the total word count for the sample. This approach is based upon the lognormal model developed by G. Herdan, which assumes that the total vocabulary is distributed normally when the logarithms of the word frequencies are used. The program also parses sentences into noun phrases, verbals, adverbals, prepositional phrases, and infinitives, according to simple rules given by Norman C. Stageberg (An Introductory English Grammar). Printouts for text include the following average numbers per sentence: words; prepositional phrases; verbal lengths; grammatical groups. Also includes the fraction of grammatical groups at the left of the first main verb to the total number of groups per sentence, as an index of left branching memory problems. Also provides: number of different sentence lengths and their frequency of occurrence; each word and the number of times it is used, grammatical use, and syllable count. In addition, the use count is divided by the total word count to yield a fractional use representation, and the printout also includes SFI values. Finally, statistics are given for average syllable count, noun-pronoun fraction of total number of words, and conjunction fraction of total number of words, and for average SFI and frequency of each SFI value. The average SFI is given for grades 3-9 based upon American Heritage data; it yields a rate of change that is linear up to approximately grade 6-7, rising rather sharply thereafter up to grade 9. In further analysis of first grade readers, the weighted SFI correlated in the .90s with all variables except "groups before verbs" and "conjunctions per sentence," including not only word variables but also such sentence variables as verbs per sentence, prepositions per sentence, and nouns per sentence. The program was applied to three bodies of text: three first grade readers, dental text (without use of the lexicon), and programmed and regular grammar texts. The program uses IBM OS/360 Assembler Language, and can analyze text employing a vocabulary of several thousand words. An interesting feature is the ability to punch new vocabulary cards for all words not in the lexicon.

Pearson, P. D. The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. Reading Research Quarterly, 1974-75, X, 155-192. (Also available under the same title, except for the omission of "certain," as an unpublished doctoral dissertation at the University of Minnesota, 1969. Because of sufficient detail in the article, it was used in place of the dissertation.)

Rewrote sets of unrelated short sentences (3-10 words in length) into forms varying in terms of: (1) presence or absence of cue (logical relationship), (2) cause-effect versus effect-cause order, and (3) one-sentence versus two-sentence form. Administered in three experiments (the first having three sub-parts) to: Experiment I, 64 third and fourth graders; Experiment II, 24 fourth graders; and Experiment III, 8 fourth graders. Used as dependent variables: I, errors, adjectival responses, clausal responses, adjective-noun, and noun responses; II, preferences for forms; and III, aided recall. Based predictions upon readability versus transformational (deep structure) versus chunk (semantic) models. Though the author felt the results were ambiguous, he indicated that virtually all of the data were explainable in terms of the chunk model, while virtually none were explainable by the readability or deep structure positions. Possible contributors to the negative readability results: rewriting (failed to use readability data that show one-sentence form using "and" conjunction to be as readable as two-sentence form); tests used (found typical error data to be non-discriminating in most cases); subjects (used only small number of third and fourth graders); administration (tested individually in two cases, raising level of motivation). Author acknowledged that use of short, related sentences might not generalize to prose context; failed to note, however, the studies showing that semantic factor (vocabulary) contributes much more heavily to effects of readability than does syntactic factor (sentence length).

Pitcher, R. W. An experimental investigation of the Flesch readability formula as related to adult materials. Unpublished doctoral dissertation, University of Michigan, 1953.

Rewrote three passages of approximately 500 words each on content characterized as familiar, abstract, or technical from original Flesch Reading Ease level of 30-40 (approximately "some college" level) to 50-60 (eighth grade level) and 70-80 (sixth grade level). Administered comprehension tests consisting of 10 true-false questions to University of Michigan students, but used only 102 subjects classified as good readers on basis of test scores of 70 percent or above for further analysis. Found significant differences in reading speed between the easier, middle, and harder levels, but also between the familiar, technical, and abstract contents (in that order). Also found significant relationships between subjects' interest ratings and performance. Possible contributors to positive results: careful rewriting with large differences in readability; used words read per minute for rate; motivation not specially raised by experimental conditions. However, comprehension differences were not measurable owing to selection of only high scorers on test; also, significance of differences in rate disappeared when ems per minute was used in place of words per minute.

Powers, R. D. A recalculation and partial validation of four adult readability formulas. Unpublished doctoral dissertation, University of Wisconsin, 1957.

New regression formulas were calculated for the 1948 Flesch Reading Ease formula, the Dale-Chall formula, and the Farr-Jenkins-Paterson simplification of the Flesch formula. The variables of words with more than two syllables, as used by Gunning, was also investigated as a possible additional simplification of the Flesch formula. The original Flesch and Dale-Chall formulas were based upon the 1926 Edition of the McCall Crabbs Standard Test Lessons in Reading; the appearance of the 1950 Edition, in which the passages and questions were found to have been changed considerably, made the recalculation possible. The following data present the original and recalculated versions of the formulas, plus their original and revised correlations (r) with the McCall-Crabbs criterion (where available), along with coefficients of determination (R^2) for the recalculated formulas and an error term consisting of two standard errors on each side of a mean value of 7.5 (grade level).

	Formula	r	R^2	Error
Dale-Chall original:	3.6365 + .0496 sl + .1579 ndw	.70		
Dale-Chall revised:	3.2672 + .0596 sl + .1155 ndw	.70	.5092	1.55
Flesch original:	206.85 - 1.015 sl - .846 w	.71		
Flesch revised:	2.2029 + .0778 sl + .0455 w	.64	.4034	1.71
Gunning original:	.4 (sl + 1 w)			
Gunning revised:	3.0680 + .0877 sl + .0948(1 w)	.59	.3440	1.80
Farr-Jenkins-Paterson original:	- 31.517 - 1.015 sl + 1.599 ms			
Farr-Jenkins-Paterson revised:	8.4335 + .0923 sl - .0648 ms	.58	.3407	1.80

(Note that the more common criterion of one standard error is well below one grade level for the revised Dale-Chall and Flesch formulas, and even slightly below for the Gunning and Farr-Jenkins-Paterson formulas.) The author concluded that: (1) the recalculated formulas should be used in preference to the original; (2) the Dale-Chall formula gained a little and the Flesch lost a little in recalculation, making the range of error smaller for the former; (3) the Dale-Chall should be used in preference to the Flesch for precision use; (4) the use of a three-or-more syllable measure works as well as a monosyllable measure as possible simplifications of Flesch's formula, but neither is as precise as the Flesch (or Dale-Chall) or as powerful for prediction. The author also provided tables for determining formula scores without the use of mathematics, as well as suggested norms for interpretation of scores.

Robinson, H. A., & Hittleman, D. R. <u>Readability of high school text passages before and after revision</u>. Final Report, Project No. I-B-025, Contract No. OEA 2-71-0025. Washington, D. C.: U. S. Office of Education, January 1973. The above study appears to be a more extensive sequel to the doctoral dissertation of D. R. Hittleman, The readability of subject matter material rewritten on the basis of students' oral reading miscues (Unpublished doctoral dissertation, Hofstra University, 1971). Since both studies gave essentially similar results, the abstract below is based upon the Robinson-Hittleman report.

Revised original passages of approximately 375 words on social studies, English and science topics, using the oral reading miscues of a stratified sample of 23 ninth grade students as the basis for the revisions. After subject-matter experts had verified the equivalency of the content of the original and revised passages, cloze versions were prepared by deleting 50 content words (to the extent possible, the same words were deleted in both original and revised versions). Administered the cloze tests to 96 ninth grade students in regular classes, with responses scored according to both exact-word and synonym methods. Results of the descriptive phase of the study indicated that (1) students, in oral reading, attempt to readjust syntactical patterns of subject-matter passages in order to gain meaning, and (2) analysis of miscues can reveal portions of passages that appear to be confusing and/or ambiguous. Results of the experimental phase of the study indicated that there was no significant difference in readability (as measured by cloze scores), between the original and rewritten passages. Possible contributors to negative results: rewriting (no comparison was made of readability grade levels of original and revised passages); test used (only one cloze version was used; also, content words only were deleted, and attempt was made to delete same words in original and revised versions); level of motivation of subjects (testing was done in regular classrooms, though not by regular instructors). However, original and rewritten versions were judged equivalent in content by subject-matter experts.

Schwimmer, S. <u>The relationship of readability to reading comprehension.</u> Unpublished doctoral dissertation, The University of Connecticut, 1971.

Rewrote passages from the Iowa Test of Basic Skills to make them less readable (more difficult). Used the READSCORE formula, but did not indicate grade levels of difficulty or word or sentence counts for the original and rewritten passages (reported only that judges found the passages to differ considerably in readability). Administered the passages and the four-choice test questions that accompany the appropriate Iowa Test passages to 242 pupils in grades 3-6, allowing 55 minutes for reading and testing. Found a strong correlation between I. Q. scores (Otis) and reading comprehension scores, but relationship between increased style difficulty and decreased comprehension was not significant. Possible contributors to negative results: rewriting (changes in sentence length between original and harder rewrite were much greater than changes in vocabulary--in some cases, word length was actually shorter in the harder version; also, used printed text in original and typewritten text in harder rewrite); test used (used original test--did not rewrite for harder version); subjects used (rather large differences in mean I. Q. scores between grades, but did not run covariance analysis); motivation level of subjects (probably raised motivational level by using passages from a regular test, and by having regular teachers test in regular classrooms, and increased effect of motivation by allowing <u>S</u>s to refer to text while answering questions under liberal time limits). However, asked 10 judges to determine that content of original and harder rewrite had not changed.

Siegel, A. I., Lautman, M. R., & Burkett, J. R. Reading grade level adjustment and auditory supplementation as techniques for increasing textual comprehensibility. Journal of Educational Psychology, 1974, 66, 895-902.

Rewrote three sets of materials: (1) a course study guide used during two intermediate weeks of a materiel facilities course; (2) a home study manual consisting of career development course guides; and (3) a technical reference manual. Used traditional readability variables: shorter words; shorter and less complex sentences; more personal writing; active rather than passive voice; step-wise presentation of material. Automated Readability Index scores were approximately five and a half grade levels easier for revised than original materials, whereas FORCAST scores were less than about one and a third grade levels different. Presented material to Ss in written form or written plus oral form. Administered course study guide materials over 2-week period, with liberal reading time, to 65 Air Force Specialist trainees, followed at end of period by judgments of difficulty and likability of material and 60-item multiple-choice test. Administered home study manual to 49 Air Force trainees, allowing 2 hours for reading, with text removed during testing on 30-item multiple-choice test. Administered reference materials to 49 Air Force trainees (same as above), allowing 20 minutes for 30-item multiple-choice test. Results showed: (1) no significant comprehension difference in all three experiments attributable to auditory supplementation; (2) significant comprehension difference attributable to readability only for home study manual; (3) significant comprehension difference for high and low reading ability groups in experiments with course study guide and reference materials, but not home study materials; and (4) revised (more readable) materials judged more likable but not more difficult. Authors explain negative results with the course study guide materials in terms of high degree of redundancy in 2-week course period owing to many other media used during that time; positive results with home study manual said to be due to lack of such redundancy. Readability expected by authors to have little practical effect for reference material. Other possible contributors to mixed results: writing in revised material (included personal style along with more typical readability variables; also, lack of change in revised version according to FORCAST formula raises questions concerning revision); tests used (no comprehension difference shown between high-ability and low-ability Ss on home study manual, raising questions concerning significant effect found for readability); administration (intended to raise level of motivation and provide liberal reading time, but text not present during testing in two parts of study); judgments (failure to find judgments of difficulty favoring revision raises questions concerning adequacy of revision). However, revised materials appeared satisfactory according to some tests (cloze test scores for samples significantly favored revised versions, as did judgments of likability); mixed results raise questions concerning the positive difference attributable to readability found with home study manual.

Thompson, E. N. Readability and accessory remarks: Factors in problem solving in arithmetic. Unpublished doctoral dissertation, Stanford University, 1967.

Wrote 10 arithmetic "word problems" at readability grade levels of 2.7 (Spache formula), and 8.7 (Dale-Chall formula), changing both word and sentence variables, also using the additional non-readability factor of accessory remarks (to give an easy or difficult impression). Administered to 368 sixth-grade pupils, either low or high in mental ability, and allowed 20 minutes for reading. Found significant main effects on problem solution for mental ability and readability, but not for accessory remarks; also found significant interaction for mental ability and readability, with improved readability helping low-ability pupils more than high-ability. Possible contributors to positive results: large differences in readability; carefully written versions; unique performance criterion; no attempt to increase motivational factors except for accessory remarks variable. However, the more readable version of problems averaged 27 words and the less readable 29 words (some of which seemed extraneous); this, plus a liberal reading time increased likelihood of positive results (which were credited entirely to readability).

Watson, P. G. Readability and comprehension of written instructional materials in homogeneous groupings of vocational/technical students in an urban community college. Unpublished doctoral dissertation, 1971.

Rewrote passage of approximately 250 words on types of communication and patient stress from 16th grade level to 11th and 7th grade levels, using the 1952 Gunning readability formula. Prepared a single cloze test form over each of the above versions, using an every-fifth deletion ratio, beginning with one of the first five words chosen randomly. Based test results on exact-word responses of 135 females in community college in 9 groups of 15 each (3 levels of readability and 3 levels of verbal intelligence, as measured by GATB); tested in regular classrooms, but administered test himself. Found both readability and verbal intelligence significant at .005. Scores on 11th and 7th grade versions were significantly higher than on 16th version; scores on 11th also somewhat higher than on 7th version, but not significantly. Rescored using synonyms instead of exact words, and found superiority of 11th and 7th versions over 16th version still greater. Effect found only for high and medium verbal intelligence group, not low. Correlations between verbal intelligence scores and cloze scores tended to increase as readability improved. Possible contributors to positive results: careful rewriting (wide range of differences); moderate level of motivation (tested in regular classroom, but regular instructor did not administer the test, and Ss were toid that the experiment was not a test and high scores were not expected). Possible contributors to mixed results (11th grade version superior to 7th): used only one cloze version (which may have been more difficult in case of former than latter) tendency for closer scores when synonyms were accepted (beginning college students might tend to use responses closer to 11th grade level than 7th grade level in 7th grade cloze test, which might penalize them in exact-word scoring); liberal reading-testing time.

Williams, D. L. The effect of rewritten science textbook materials on the reading ability of sixth-grade pupils. Unpublished doctoral dissertation, University of Illinois, 1964.

Rewrote a 4,879-word science chapter from original sixth-grade level to third-grade level, changing both sentence and work variables to improve readability, but without altering technical terms. Administered sections of original and revised versions over 3-day period, followed by multiple-choice tests (number of items totalled 129), to 417 sixth-grade pupils (effective N was 322). Found significantly higher reading rate (on timed reading) and comprehension on third than sixth-grade level passages. Results held for below-average, average, and above-average readers. Possible contributors to positive results: careful rewriting; long, well-developed test; allowed only one reading, with text not present during testing; no emphasis on speed or other special emphasis on motivation. Testing by regular classroom teachers in regular classrooms may have raised motivation somewhat; on the other hand, clarifying technical vocabulary in rewritten version, not normally a part of improving readability, may have contributed to positive results.

ANNOTATED BIBLIOGRAPHY:

SELECTED BOOKS, ARTICLES, AND REPORTS
DEALING WITH READABILITY/COMPREHENSIBILITY

ANNOTATED BIBLIOGRAPHY: SELECTED BOOKS, ARTICLES AND REPORTS DEALING WITH READABILITY/COMPREHENSIBILITY

American Institutes for Research in the Behavioral Sciences. Word frequencies in technical instruction. Report for Contract NOP-1498, AIR-D53-8/64-RP, August 1964.

Presents tables of word frequencies from 60 samples of lectures, training manuals, and instructional responses of military trainees. A master word list was constructed, in which every word appearing in any of the samples is listed alphabetically. A second listing, the "Principal Word List," contains a rank ordering, by frequency, of words in the master list, with part of speech, meaning, number of samples in which the word occurred, and the total number of appearances in each of the three types of samples. Only words which occurred at least 60 times across the 60 samples are included in the Principal Word List. The list includes approximately 1,745 words.

Bormuth, J. R. Development of readability analyses. Final report to U. S. Department of Health, Education, and Welfare, Office of Education, Bureau of Research, Contract No. OEC-3-7-070052-0326, March 1969.

An extensive correlational study which examines the relationship of a vast number of variables to readability. Intended as a means of determining how to increase the effectiveness with which students acquire knowledge from written instructional materials. Examines the difference between those variables which are only correlated with readability and those which have a causal effect on comprehension. Examines the relationship of a large number of psycholinguistic variables to readability, in addition to such variables as word and sentence length. Examines the "cloze test" in detail with regard to its relationship to readability. Presents a number of readability formulas for various purposes using cloze scores as criteria. A major problem in using these formulas is the difficulty of comparing cloze scores to reading grade level. Gives a clear and comprehensive statement of the implications of this study for readability research in general.

Carver, R. P. Improving reading comprehension: Measuring readability. American Institutes for Research, Final Report, Contract N00014-72-C-0240, May 1974.

Presents a standardized method which can be used to automatically convert prose training material into a form which forces trainees to read the material with at least a minimal level of comprehension ("Programmed Prose"). Concludes that this method facilitates learning under certain specified conditions (e.g., when individuals are not always highly motivated to learn, or when attention to the reading task wanes). Also presents a new method for measuring prose difficulty, the "Rauding Scale," and found it to correlate higher with the actual grade level where curriculum passages were used in school than did six other readability measures. It should be noted, however, that only 10 out of 60 students (from graduate reading courses) who were administered the "Rauding Scale Qualification Test" actually passed the test.

Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. Methodologies for determining reading requirements of military occupational specialties. Human Resources Research Organization Technical Report 73-5, March, 1973.

Describes research performed in HumRRO's Work Unit READNEED directed toward (1) refinement of methods used in earlier research in the area of MOS reading requirements, and (2) development of more general, less expensive methods than those previously used. Describes the features of the FORCAST readability formula and its application to Army technical materials. Compares the "cloze" technique to other tests in its ability to accurately measure comprehension.

Coke, E. U., & Rothkopf, E. Z. Note on a simple algorithm for a computer-based Reading Ease Score. Journal of Applied Psychology, 1970, 54, 208-210.

Presents a method of estimating syllable count for the computation of the Flesch Reading Ease measure of readability. Such a method would be of considerable value due to the difficulty of counting word syllables by computer. Found a Pearson product-moment correlation of .92 for a machine count of vowels per word with a man-made count of syllables. Applying the vowel count to 200 successive 100-word samples from a physics textbook, and using the resulting statistics as estimates of the true parameters of the passages, developed a table showing the probability of misclassification of a passage by five or more points on the Reading Ease scale for a large number of different sample sizes.

Department of the Air Force, Air Force Pamphlet 10-1, Guide for Air Force Writing, July 1969.

Provides the standard writing guide for all Air Force people who write or approve the writing of other people. Covers the principles of clear, logical thinking and writing, with specific suggestions for applying those principles to the most common types of Air Force written communication. Deals primarily with attitudes and practical methods and techniques, rather than with theories and rules of grammar, punctuation, and composition. Easy-to-read manual, written mainly in the second person. Contains well-written guidelines for application of the Fog Count in measuring the readability of materials.

Federman, P. J., MacPherson, D. H., & Siegal, A. I. Normative development for submarine sonar manuals. January 1970, Applied Psychological Services, Inc., Contract N00014-67-C-0450, Sonar Technology Division, Naval Ships Systems Command, Engineering Psychology Branch, Psychological Sciences Division, Office of Naval Research.

Presents a series of norms for use in evaluating the readability/comprehensibility of submarine sonar manuals, compiled through the use of the cloze technique. Compares the manual for a new sonar system with 15 manuals already in use, with technicians of four levels of experience serving as subjects. Found the new manual to be significantly inferior to other manuals and presents a range of acceptability for the 15 manuals with which it was compared. Found a statistically significant difference between pay grades in terms of readability/comprehensibility.

Flesch, R. How to write, speak, and think more effectively. New York: Harper & Brothers, 1960.

A compilation of material from six earlier books which, in the authors' view, provides a step-by-step sequence of lessons designed to give one a different set of mental habits. The book is replete with examples of "good" and "bad" writing, speaking, and thinking, and offers many exercises in improving one's skills in these areas. The final part of the book presents explicit guidelines for testing and raising the readability of materials. While it applies primarily to "lay" writing, it includes frequent reference to more technical applications. A unique feature would seem to be its attempt to apply the philosophy of science to improvement in writing, reading, and speaking skills.

Gunning, R. The technique of clear writing. (Rev. ed.) New York: McGraw-Hill Book Company, 1968.

Tends to be more "prescriptive" in approach than the earlier (1952) edition, which was more heavily oriented toward "description" (prediction) of readable writing. With regard to measuring reading difficulty, the book emphasizes that the Fog Index (and readability formulas in general) are tools, not rules. Formulas are warning systems, not formulas for writing. Based in large part on the proposition that there are definite principles of clear writing, and that one can improve both the form and the force of one's own writing by building these principles into his own personal style. Outlines 10 principles of clear writing, pointing out that the one most often violated is that against complexity. Offers a chapter on technical writing, which, although general, contains valuable hints on the special problems inherent in such writing.

Harris, A. J., & Jacobson, M. D. Revised Harris-Jacobson readability formulas. Paper presented at the College Reading Association convention, Bethesda, Maryland, October 1974.

For each of seven primary grade levels, examined 10 primer passages taken from each of six basal readers. From this large number of words, constructed three separate lists containing words common to particular grade levels. The purpose for this work was to use as a variable in readability formulas the "percent of uncommon words" (words not on a given list). The "Short List" (containing 2,792 words) was used in computing regression equations along with average sentence length, percent of long words (six or more letters), mean number of letters per word, and "spelling patterns." Five readability formulas for special purposes are reported. Found that spelling patterns were the best single measure of vocabulary difficulty, which is considered the most valid single indicator of readability. A special computer program developed by Jacobson is necessary in order to apply formulas using this variable. Its high correlations with readability (12 patterns correlating N .92) indicate that its use, although complicated, would be well worthwhile.

Hicks, T. G. Writing for engineering and science. New York: McGraw-Hill 1961.

Gives a basic and comprehensive coverage of the field of modern technical writing, addressing all major forms of technical writing used today--reports, articles, papers, manuals, and specifications. The first five chapters cover general methods of writing--outlines, grammar, usage, illustrations, and tables. The next 10 chapters discuss important writing tasks, such as engineering and scientific reports, technical and scientific articles, instruction manuals and bulletins, military manuals, specifications, sales, and news writing. Presents a five-step approach to training technical writers, including instructive exercises in realistic writing situations.

Joyce, R. P., & Chenzoff, A. P. Improving job performance aids through condensation, dual-level presentation, promotion of learning, and entry by malfunction symptoms. Air Force Human Resources Laboratory, AFHRL-TR-74-12, March 1974.

Aimed at identifying factors which would reduce the bulk of JPAs and at defining techniques which would make the typical JPA usable by both trained and less experienced technicians. Develops several techniques for arrangement of text and illustration formats which lead to greater flexibility and greater space utilization. Suggests an approach for the development of "fully proceduralized JPAs," which would not only provide explicit directions for troubleshooting, but would at the same time teach a man to troubleshoot.

Kern, R. P., Sticht, T. G., Welty, D., & Hauke, R. N. Guidebook for the development of Army training literature. Human Resources Research Organization for U. S. Army Research Institute for the Behavioral and Social Sciences, Contract DAHC 19-73-C-0051, April 1974.

Points out that most current Army training literature is too difficult for many Army readers and oriented toward content area topics rather than job performance. In an attempt to remedy this situation, the guidebook was developed as a "job aid" for the writer as opposed to a study text. Concentrates on providing many examples of good and bad writing for writers to use as models. Gives a good deal of attention to the reading abilities of Army enlisted personnel and how the writer can use this information in his work. Emphasizes the need for "performance oriented" writing--telling the user what to do and how to do it. Gives detailed instructions for applying the FORCAST readability formula to material to determine its reading grade level. Offers a brief but intuitively reasonable chapter on the use of illustrations in technical writing.

Kincaid, J. P. Making technical writing readable (the Fog Count and an alternative). Human Factors Society Bulletin, Vol. XV, No. 5, May 1972.

Describes the development of the Automated Readability Index (ARI) as the Air Force's alternative to the traditional Flesch count and the Fog Count. Compares the Fog Count and the ARI for both test-retest and between subject reliability, finding coefficients of about .50 for the Fog and approaching 1.0 for the ARI. Reports the 95% sampling confidence interval for the ARI to be about one-half of one grade level. Recommends the use of the ARI as an alternative to the Fog Count for USAF technical writing based on its reliability, validity, and ease of usage.

Klare, G. R. The measurement of readability. Ames, Iowa: Iowa State University Press, 1963.

Offered primarily as a review of research literature in the field of readability. Critically reviews past research and considers future directions that readability research may profitably take. Discusses in detail the many variables that must be considered in producing readable writing and in conducting readability research. Discusses the advantages and limitations of readability formulas in general and offers suggestions for deciding which of the many formulas to use in a given situation. Presents an extensive annotated bibliography divided into sections dealing with readability in general, formulas, alternatives to formulas in measuring readability, reliability and validity, among others. It is basic enough to be easily readable and yet covers the field exhaustively.

Klare, G. R. Analysis of the readability level of selected United States Armed Forces Institute printed instructional materials. Report to the United States Armed Forces Institute, Madison, Wisconsin, Contract No. DAHC11-71-C-0166, September 1971.

Reports a study dealing with the relationship between readability of instructional materials and students who begin correspondence courses but do not complete them. Analyzed four high school, twenty college, and six occupational courses for readability using an automated version of the Flesch "Reading Ease" formula. Reports those courses deemed too difficult for the particular audience. Makes suggestions for possible experimental changes in the materials for the purpose of increasing student completion rate.

Klare, G. R. The readability of printed instructional materials: Further analysis and revision. Final Report to the United States Armed Forces Institute, Madison, Wisconsin, January 1973.

Analyzes three USAFI correspondence courses for readability and compares these indices with percentage of students completing the course. Discusses the project in terms of three aspects: prediction of readability, production of readable material, and automation of readability analyses. Central to the project was the evaluation of the CARET I computer system in terms of rewriting of unsatisfactory (i.e., difficult to read) materials.

Klare, G. R. Assessing Readability. Reading Research Quarterly, Vol. X, No. 1, 1974-1975.

Reviews measurement and predictive devices in the area of readability which have been developed since the author's earlier review of such techniques prior to 1960. Contains suggestions for selecting a formula considering specific need, complexity, manual versus machine application, use of word length versus word lists, and use of sentence length versus sentence complexity. Compares the validity of formulas and reader judgments in determining readability.

Klare, G. R. Personal Communication, February, 1975.

The author reviewed 15 source books giving suggestions for clear writing. One hundred and fifty-six different suggestions were identified and grouped into three categories: high agreement, low agreement, and disagreement or contradiction. Even in the "high agreement" category, the range of mention of suggestions was only three to six of the 15 books. In the "low agreement" category, certain intuitively obvious suggestions were mentioned in either one or two of the books. And there was outright disagreement with regard to such variables as paragraph length, use of repetition, and conciseness.

Klare, G. R., Rowe, P. P., St.John, M. G., & Stolurow, L. M. Technical Report No. B.2.1, A first version of a computer-aided revising, editing, and translating system (CARET I). Cambridge: Harvard Computer-Aided Instruction Laboratory, 1969.

Describes a preliminary version of a system, called CARET I, which provides the following categories of aid for judging the suitability of a piece of written material and rewriting it if necessary: (1) computer application of readability formulas; (2) computer formatting of material to make editing and revision more convenient; and (3) suggestions from the research literature intended to help the writer or teacher make changes in material so that it will be more readable.

Klare, G. R., Sinaiko, H. W., & Stolurow, L. M. Cloze procedure: A convenient readability test for training materials and translations. International Review of Applied Psychology, 1972, 21(2), 77-106.

Examines the cloze procedure as an easy, fast, and inexpensive way of comparing the intelligibility of different versions of instructional material. Differentiates the "predictive" nature of readability formulas from the "measurement" nature of the cloze procedure. Provides direction for the construction and administration of cloze tests, such as the selection of test passages, the preparation of cloze versions, selecting of subjects, scoring responses, and analyzing the results. Presents results of experiments applying cloze procedures to training materials and to translations of technical English. Presents an extensive bibliography to 1971.

Krohn, C. A., Clayton, C. A., & Dunn, J. W. Evaluation of SIMMS for AN/SRC-20 radio set. Naval Electronics Laboratory Center Contract No. N00123-71-D-0168, Research Triangle Institute, November 1971.

Reports results of research comparing two manuals, SIMMS and conventional, in terms of resulting level of troubleshooting ability of equally trained and experienced technicians. Findings suggest that SIMMS is a superior approach for technicians with little training in terms of system troubleshooting, course examinations, and assembly identification. Technicians with moderate training tend to perform equally well with either type of manual, while those with high training tend to perform better with conventional manuals. Results are only suggestive due to interaction of manual and training site, small numbers of subjects in some categories, and small practical, although statistically significant, differences in performance scores.

Post, T. J., & Price, H. E. Requirements and criteria for improving reading comprehension of technical manuals. BioTechnology, Inc., Naval Sea Systems Command, Technical Publications Branch Contract No. N00024-74-C-5426, November 1974.

Report points out that a significant "gap" exists between the reading comprehension level of Navy technical manuals and the reading abilities of new Navy technicians. Gives requirements and criteria to improve the readability of manuals which (1) can be used either in original preparation of manuals or to determine if an existing manual needs re-writing; (2) address organization and communication variables as well as nonprose presentations; and (3) require testing to determine their usability and acceptability for Navy technicians.

Siegel, A. I., & Burkett, J. R. Application of structure-of-intellect and psycholinguistic concepts to reading comprehensibility measurement. Air Force Human Resources Laboratory, Technical Training Division, Lowry Air Force Base Technical Report, AFHRL-TR-74-49, September 1974.

Compared two sets of reading materials, one loaded heavily on variables drawn from psycholinguistics and intellective function literature and the other loaded lightly on these variables. Demonstrated the effects of these variables on comprehensibility of the material. Defined methods for improving the readability/comprehensibility of textual materials and discussed methods for computer analysis of text using the variables from this experiment.

Siegel, A. I., Lambert, J. V., & Burkett, J. R. Techniques for making written material more readable/comprehensible. Air Force Human Resources Laboratory Report AFHRL-TR-74-47, August 1974.

Attempts to apply the findings of modern scientific research relative to how certain kinds of written materials impose a heavy load on a reader's mind. Tells a writer, on the basis of these findings, how to avoid placing a heavy mental load on his reader and thus, at the same time, make his written material more understandable. Presents 12 techniques for increasing readability/comprehensibility, including reduction in morpheme count, sentence voice, and sentence tense. Suggests five methods for measuring readability, such as the use test, the cloze procedure, and expert ratings. Purposely avoids discussion of readability "formulas."

Sticht, T. G., Caylor, J. S., Kern, R. P., & Fox, L. C. *Determination of literacy skill requirements in four military occupational specialties*. Human Resources Research Organization Technical Report 71-23, November 1971.

Describes results of research on the extent of usage of job printed materials and job listening sources as a function of the reading difficulty level of the materials and the reading ability of Army job incumbents. Discusses methods for reducing discrepancies between personnel literacy skill levels and the literacy demands of the job by remedial literacy training or redesign of job literacy materials. Presents research findings in the light of the implications for selection, training, and research.

Whalen, G. V. *Study for investigation of the application of the Readability Ease Assessment Device to Navy technical publications*. A Research Study Description submitted to U. S. Navy Material Command, Washington, D. C., February 1974.

The purpose of the proposed study is to determine how well the Readability Ease Assessment Device (READ) can be applied to the appraisal of existing publications and how it might best be utilized in the production of Navy technical materials. Sample materials consisting of 20 randomly selected pages from each of 10 technical areas will be measured by READ and Flesch indices. The methods will then be compared for apparent accuracy, reliability, efficiency, time saving, and general effectiveness. An optimal approach for Navy utilization of READ with regard to the number of devices, organizational locations, and techniques for usage will be recommended.

BIBLIOGRAPHY OF OTHER SELECTED READABILITY/COMPREHENSIBILITY SOURCES

BIBLIOGRAPHY OF OTHER SELECTED READABILITY/COMPREHENSIBILITY SOURCES

I. SELECTED READABILITY/COMPREHENSIBILITY BIBILOGRAPHY

American Institutes for Research in the Behavioral Sciences. Word frequencies in technical instruction. Report for Bureau of Naval Personnel, Department of the Navy and Defense Language Institute, Department of Defense, Contract NOp-1498, AIR-D53-8/64-RP, August 1964.

Booher, H. R. Effects of still pictures and printed words on compre hension of job performance instructions. Unpublished doctoral dissertation, The Catholic University of America, 1973. (Also available from University Microfilms, Document 73-12, 463.)

Bormuth, J. R. Cloze test readability: Criterion reference scores. Journal of Educational Measurement, 1968, 5, 189-196.

Bormuth, J. R. The implications of the use of cloze procedure in the evaluation of instructional programs. Occasional Report No. 3, Center for the Study of Evaluation of Instructional Programs, University of California, Los Angeles, 1967. (a)

Bormuth, J. R. Comparable cloze and multiple choice comprehension test scores. Journal of Reading, 1967, 10, 291-299. (b)

Bormuth, J. R. Readability: A new approach. Reading Research Quarterly, 1966, 1, 79-131.

Bormuth, J. R. Development of readability analysis. Final Report, Project No. 7-0052, Contract No. OEC-3-7-070052-0326. Bureau of Research, U. S. Office of Education, 1969.

Brown, W. R. Science textbook selection and the Dale-Chall formula. School Science and Mathematics, February 1965, 65, 164-167.

Carver, R. P. New techniques for measuring and improving reading com prehension. American Institutes for Research, Techbical Report No. 1, Contract N00014-72-C-0240 for Personnel and Training Research Programs, Office of Naval Research, February 1973.

Carver, R. P. Improving reading comprehension: Measuring readability. American Institutes for Research, Final Report, Contract N00014-72-C-0240 for Personnel and Training Research Programs, Office of Naval Research, May 1974.

Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. Methodologies for determining reading requirements of Military Occupational Specialties. Human Resources Research Organization Technical Report 73-5, March 1973.

Chall, J. S. Readability: An appraisal of research and application. Bureau of Educational Research Monographs, No. 34, Ohio State University, 1958.

Coke, E. U., & Rothkopf, E. Z. Note on a simple algorithm for a computer-based Reading Ease score. Journal of Applied Psychology, 1970, 54, 208-210.

Coleman, E. B. Improving comprehensibility by shortening sentences. Journal of Applied Psychology, 1962, 46, 131-134.

Coleman, E. B. The comprehensibility of several grammatical transformations. Journal of Applied Psychology, 1964, 48, 186-190.

Coleman, E. B. Experimental studies of readability. In J. R. Bormuth (Ed.) Readability in 1968. Champaign, Illinois: National Council of Teachers of English, 1968.

Coleman, E. B. Developing a technology of written instruction: Some determiners of the complexity of prose. In E. Z. Rothkopf and P. E. Johnson (Eds.) Verbal learning research and the technology of written instruction. New York: Teachers College Press, Columbia University, 1971.

Condon, C. F. M. Survey of Navy Technical Manual Systems and procedures. Bradford Computer & Systems, Inc. for Navy Ship Research and Development Center, Contract N00600-75-D-0152, April 1975.

Dale, E., & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 37-54.

Department of the Air Force, Air Force Pamphlet 10-1, Guide for Air Force writing. July 1969.

Duffy, T. M., Carter, J. F., Fletcher, J. D., & Aiken, E. G. Language skills: A research and development plan for the Naval service. San Diego, California: Navy Personnel Research and Development Center Special Report. (in press)

Farr, J. N., & Jenkins, J. J. Tables for use with the Flesch readability formulas. Journal of Applied Psychology, 1949, 33, 275-278.

Farr, J. N., Jenkins, J. J., & Paterson, D. G. Simplification of Flesch Reading Ease Formula. Journal of Applied Psychology, 1951, 35, 333-357.

Federman, P. J., MacPherson, D. H., & Siegal, A. I. Normative development for submarine sonar manuals. Applied Psychological Services, Contract N00014-67-C-0450, for Sonar Technology Division, Naval Ships System Command, Engineering Psychology Branch, Office of Naval Research, January 1970.

Flesch, R. F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Flesch, R. F. How to test readability. New York: Harper & Brothers, 1950.

Flesch, R. F. How to write, speak, and think more effectively. New York: Harper & Brothers, 1960.

Fry, E. B. A readability formula that saves time. Journal of Reading, 1968, 11, 513-516.

Gunning, R. The technique of clear writing. (Rev. ed.) New York: McGraw-Hill, 1968.

Guthrie, J. T. Learnability versus readability of texts. Journal of Educational Research, 1972, 6, 273-280.

Harris, A. J., & Jacobson, M. D. Basic vocabulary for beginning reading. The Reading Teacher, January 1973, 392-395.

Harris, A. J., & Jacobson, M. D. Some comparisons between Basic Elementary Reading Vocabularies and other word lists. Reading Research Quarterly, 1973-1974, No. 1, 87-109.

Harris, A. J., & Jacobson, M. D. Revised Harris-Jacobson readability formulas. Paper presented at convention of the College Reading Association, Bethesda, Maryland, October 1974.

Harris, A. J., & Jacobson, M. D. Comparative predictions of tested reading comprehension of intermediate and secondary pupils. Paper presented at annual convention of American Educational Research Association, Washington, D. C., April 1975.

Jacobson, M. D. Reading difficulty of physics and chemistry textbooks. Educational and Psychological Measurement, 1965, XIV, No. 2., 449-457.

Jacobson, M. D. Predicting reading difficulty from spelling. Spelling Progress Bulletin, Spring 1974, 14, 8-10.

Joyce, R. P., & Chenzoff, A. P. Improving job performance aids through condensation, dual-level presentation, promotion of learning, and entry by malfunction symptoms. Air Force Human Resources Laboratory, AFHRL-TR-74-12, March 1974.

Kern, R. P., Sticht, T. G., & Fox, L. C. Readability, reading ability, and readership. HumRRO Professional Paper 17-70. Alexandria, Virginia: Human Resources Research Organization, June 1970.

Kern, R. P., Sticht, T. G., Welty, D., and Hauke, R. N. Guidebook for the development of Army training literature. Human Resources Research Organization for U. S. Army Research Institute for the Behavioral and Social Sciences. Contract No. DAHC 19-73-C-0051, April 1974.

Kincaid, J. P. Making technical writing readable (the Fog Count and an alternative). Human Factors Society Bulletin, Vol. XV, No. 5, May 1972.

Kincaid, J. P., Yasutake, J. Y., & Geiselhart, R. Use of the Automated Readability Index to assess comprehensibility of Air Force technical orders. SEC-TR-67-47. Wright-Patterson Air Force Base, Ohio: Aeronautical Systems Division, 1967.

Kincaid, J. P., & Delionbach, L. J. Validation of the Automated Readability Index: A follow-up. Human Factors, 1973, 15, 17-20.

Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel. Naval Technical Training Command, Research Branch Report 8-75, February 1975.

Klare, G. R. The measurement of readability. Ames Iowa: Iowa State University Press, 1963.

Klare, G. R. The role of word frequency in readability. In J. R. Bormuth (Ed.), Readability in 1968. Champaign, Illinois: National Council of Teachers of English, 1968.

Klare, G. R. Analysis of the readability level of selected United States Armed Forces Institute printed instructional materials. Report to the United States Armed Forces Institute, Madison, Wisconsin, Contract No. DAHC11-71-C-166, September 1971.

Klare, G. R. The readability of printed instructional materials: Further analysis and revision. Final Report to the United States Armed Forces Institute, Madison, Wisconsin, January 1973. (a)

Klare, G. R. A manual for readable writing. Draft Copy, July 1973. (b)

Klare, G. R. Assessing readability. Reading Research Quarterly, Vol. X, No. 1, 1974-1975.

Klare, G. R., Rowe, P. P., St.John, M. G., & Stolurow, L. M. A first version of a computer-aided revising, editing, and translating system (CARET I). Harvard Computer-Aided Instruction Laboratory, Technical Report No. B.2.1, 1969.

Klare, G. R., Sinaiko, H. W., & Stolurow, L. M. Cloze procedure: A convenient readability test for training materials and translations. International Review of Applied Psychology, 1972, 21(2), 77-106.

Krohn, C. A., Clayton, C. A., & Dunn, J. W. Evaluatior of SIMMS for AN/SRC-20 radio set. San Diego, California: Naval Electronics Laboratory Center Contract No. N00123-71-D-0168, Research Triangle Institute, November 1971.

Lorge, I. Predicting readability. Teachers College Record, 1944, 45, 404-409.

Miller, G. R., & Coleman, E. B. A set of thirty-six prose passages calibrated for complexity. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 851-854.

NAVSHIPS 92154.2(A). Operator's Handbook for sonar set AN/BQS-2, February 1954.

NAVSHIPS 92154.3(A). Maintenance handbook for sonar set AN/BQS-2, February 1954.

Post, T. J., & Price, H. E. Development of innovative job aid concepts. BioTechnology, Inc. for Naval Air Systems Command, Contract N00600-71-C-0835, May 1972.

Post, T. J., & Price, H. E. Requirements and criteria for improving reading comprehension of technical manuals. BioTechnology, Inc. for Naval Sea Systems Command, Technical Publications Branch, Contract N00024-74-C-5426, November 1974.

Powers, R. D., Sumner, W. A., & Kearl, B. E. A recalculation of four adult readability formulas. Journal of Educational Psychology, 1958, 49, 99-105.

Price, H. E., Post, T. J., & Kilsrud, G. Development of information measurement techniques for quality assurance of Navy aircraft maintenance job aids. BioTechnology, Inc. for Naval Air Systems Command Naval Air Development Center, Warminster, Contract N62269-70-C-0395, June 1971.

Rankin, E. F., & Culhane, J. W. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13, 193-198.

Rowan, T. C. Improving DOD maintenance through better performance aids. Information Concepts, Inc. for Advanced Research Projects Agency, Contract DAHC 15-70-0257, March 1973.

Siegel, A. I., Federman, P. J., & Burkett, J. R. Increasing and evaluating the readability of Air Force written materials. Air Force Human Resources Laboratory, Technical Training Division, Lowry Air Force Base, AFHRL-TR-74-28, August 1974.

Siegel, A. I., Lambert, J. V., & Burkett, J. R. Techniques for making written material more readable/comprehensible. Air Force Human Resources Laboratory Report AFHRL-TR-74-47, August 1974.

Siegel, A. I. & Burkett, J. R. Application of structure-of-intellect and psycholinguistic concepts to reading comprehensibility measurement. Air Force Human Resources Laboratory, Technical Training Division, Lowry Air Force Base, Technical Report, AFHRL-TR-74-49, September 1974.

Smith, E. A., & Senter, R. J. Automated Readability Index. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Division, AMRL-TR-66-220, November 1970.

Smith, E. A., Kincaid, J. P. Derivation and validation of the Automated Readability Index for use with technical materials. Human Factors, 1970, 12, 457-464.

Sticht, T. G., Caylor, J. S., Kern, R. P., & Fox, L. C. Determination of literacy skill requirements in four Military Occupational Specialties. Human Resources Research Organization Technical Report 71-23, November 1971.

Stocker, L. P. Increasing the precision of the Dale-Chall readability formula. Reading Improvement, Winter, 1971-1972, 8, 87-89.

Taylor, W. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Taylor, W. L. Recent developments in the use of "cloze procedure." Journalism Quarterly, 1956, 33, 42-46, 99.

Taylor, W. L. "Cloze" readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 41, 19-26.

Wang, M. D. The role of syntactic complexity as a determiner of comprehensibility. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 398-404.

Williams, A. R., Jr., Siegel, A. I., & Burkett, J. R. Readability of textual material: A survey of the literature. Air Force Human Resources Laboratory, Air Force Systems Command, AFHRL-TR-74-29, July 1974.

II. SELECTED COMPREHENSION BIBLIOGRAPHY

Athey, I. J. Language models and reading. Reading Research Quarterly, 1971, Vol. 8(2), 16-110.

Bormuth, J. R. Mean word depth as a predictor of comprehension difficulty. California Journal of Educational Research, 1964, 15, 226-231.

Bransford, J. D. & Johnson, M. K. Considerations of some problems of comprehension. In W. G. Chase (Ed.), Visual information-processing. New York: Academic Press, 1973, 383-438.

Chomsky, N. A. Aspects of a theory of syntax. Cambridge, Mass.: M.I.T. Press, 1965.

Chomsky, N. A., & Miller, G. A. Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.) Handbook of mathematical psychology, Vol. 2. New York: Wiley, 1963, pp. 269-321.

Coleman, E. B. Learning of prose written in four grammatical transformations. Journal of Applied Psychology, 1965, 49, 332-341.

Geyer, John J. Comprehension and partial models related to the reading process. Reading Research Quarterly, 1972, Vol. 7(4), 541-587.

Goodman, K. S. A linguistic study of cues and miscues in reading. Elementary English, 1965, 42, 639-643.

Goodman, K. S. (Ed.) The psycholinguistic nature of the reading process. Detroit: Wayne State University Press, 1968.

Goodman, K. S. Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Del.: International Reading Association, 1970, pp. 259-272.

Gough, P. B. The verification of sentences: The effects of delay of evidence and sentence length. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 492-496.

Hamilton, H. W., & Deese, J. Comprehensibility and subject-verb relations in complex sentences. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 163-170.

Hochberg, J. Attention in perception and reading. In F. A. Young & D. B. Lindsley (Eds.), Early experience and visual information processing in perceptual and reading disorders. Washington, D. C.: National Academy of Sciences, 1970, pp. 219-224.

Levin, H., Grossman, J., Kaplan, E., & Yang, R. Constraints and eye-voice span in right and left embedded sentence. Language and Speech, 1970, 15(1), January-March, 1972, 30-39.

Levin H., & Kaplan, E. L. Grammatical structure and reading. In H. Levin & J. P. Williams (Eds.) Basic studies on reading, New York: Basic Books, 1970, pp. 119-133.

Mackworth, J. F. Some models of the reading process: Learners and skilled readers. Reading Research Quarterly, 1972, Vol. 7(4), 701-733.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1963, 4, 748-762.

Miller, G. A. Some psychological studies of grammar. American Psychologist, November 1962, 11, 748-762.

Miller, G. A. The psychology of communication: Seven essays. Baltimore: Penguin Books, 1967.

Miller, G. A. & Chomsky, N. A. Finitary models of language user. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology. New York: Wiley, 1963.

Miller, G. A. & McKean, K. A chronometric study of some relations between sentences. Quarterly Journal of Experimental Psychology, November 1964, 16(4), 297-308.

Oakan, R., Wiener, M., & Cromer, W. Identification, organization, and reading comprehension for good and poor readers. Journal of Educational Psychology, 1971, Vol. 62, No. 1, 71-78.

Pearson, P. D., The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. Reading Research Quarterly, 1974-75, 2, 155-192.

Ruddell, R. B. Psycholinguistic implications for a system of communications model. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and process of reading. Newark, Del.: International Reading Association, 1970, pp. 239-258.

Simons, H. D. Reading comprehension: The need for a new perspective. Reading Research Quarterly, Vol. 4(3), 1971, 338-363.

Smith, F. Understanding reading: A psycholinguistic analysis of reading and learning to read. New York: Holt, Rinehart, & Winston, 1971.

Steiner, R., & Weiner, M. Comprehension training and identification for poor and good readers. Journal of Educational Psychology, 1971, Vol. 62, No. 6, 506-513.

Thorndike, E. L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, June 1917, 8, 323-332.

Wang, M. D. The role of syntactic complexity as a determiner of comprehensibility. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 398-404. (b)

Williams, J. P. Learning to read: A review of theories and models. Reading Research Quarterly, 1973, 7(2), 121-146.

Yngve, V. The depth hypothesis. In R. Jakobson (Ed.), Structure of language and its mathematical aspects. Providence, R. I.: American Mathematical Society, 1961, pp. 130-138.

III. SELECTED GRAPHICS BIBLIOGRAPHY

Auerbach, Inc. Auerbach on optical character recognition. New York: Auerbach Publishers, 1971.

Birkhoff, G. D. Aesthetic Measure. Cambridge, Mass.: Harvard University Press, 1933.

Evans, T. G. Descriptive pattern analysis techniques. In A. Grasselli (Ed.), Automatic interpretation and classification of images. New York: Academic Press, 1969, 79-95.

Feigenbaum, E. A. Information processing and memory. In D. A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970, 451-469.

Fetter, W. A. Computer graphics in communication. New York: McGraw-Hill, 1965.

Horn, S. Light pen input to a design package in current production. In R. D. Murray (Ed.), Computer handling of graphical information. Washington, D. C.: Society of Photographic Scientists and Engineers, 1970, 41-52.

Huggins, W. H., & Entwisle, D. R. Iconic communication: An annotated bibliography. Baltimore, Md.: The Johns Hopkins University Press, 1974.

Joyce, R. P., & Chenzoff, A. P. Improving job performance aids through condensation, dual-level presentation, promotion of learning, and entry by malfunction symptoms. Air Force Human Resources Laboratory, AFHRL-TR-74-12, March 1974.

Kern, R. P., Sticht, T. G., Welty, D., & Hauke, R. N. Guidebook for the development of Army training literature. U. S. Army Research Institute DAHC 19-73-C-0051, April 1974.

Machover, C. Computer graphics in the United States. In R. Parslow, R. Prowse, & R. Green (Eds.), Computer graphics. New York: Plenum Press, 1969, 61-84.

Matsa, S. M. What has computer graphics to offer? In R. Parslow, R. Prowse, & R. Green (Eds.), Computer graphics. New York: Plenum Press, 1969, 1-6.

Murray, R. D. (Ed.), Computer handling of graphical information. Washington, D. C.: Society of Photographic Scientists and Engineers, 1970.

Narasimhan, R. On the description, generation, and recognition of classes of pictures. In A. Grasselli (Ed.), Automatic interpretation and classification of images. New York: Academic Press, 1969, 1-42.

Naval Air Systems Command. MIARS controlled tech data with microfilm and TRUMP mechanized update techniques. Working Paper, October 1974.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Post, T. J., & Price, H. E. Requirements and criteria for improving reading comprehension of technical manuals. BioTechnology, Inc. for Naval Sea Systems Command, Technical Publications Branch, 0463, Contract N00024-74-C-5426, November 1974.

Rosenfeld, A. Picture processing by computer. Computing Surveys, 1969, 1(3), 147-174.

Rosenfeld, A. Progress in picture processing: 1969-1971. Computing Surveys, 1973, 5(2), 81-104.

Whetstone, A. Free-hand data input. In R. D. Murray (Ed.), Computer handling of graphical information. Washington, D. C.: Society of Photographic Scientists and Engineers, 1970, 11-28.

IV. SELECTED STYLE GUIDE BIBLIOGRAPHY

Comer, D. B., & Spillman, R. R. Modern technical and industrial reports. New York: G. P. Putnam & Sons, 1962.

Department of the Air Force. Air Force Pamphlet 10-1, Guide for Air Force Writing. Washington, D. C.: United States Government Printing Office, 1969.

Flesch, R. The art of readable writing. New York: Harper & Brothers Publishers, 1949.

Gunning, R. The technique of clear writing. (Rev. ed.), New York: McGraw-Hill Book Company, 1968.

Hicks, T. G. Writing for engineering and science. New York: McGraw-Hill Book Company, 1961.

Kern, R. P., Sticht, T. G., Welty, D., & Hauke, R. N. Guidebook for the development of Army training literature. Human Resources Research Organization (Western Division) for U. S. Army Research Institute for the Behavioral and Social Sciences. Contract No. DAHC 19-73-C-0051, April 1974.

Marder, D. The craft of technical writing. New York: The Macmillan Company, 1960.

Morris, J. E. Principles of scientific and technical writing. New York: McGraw-Hill Book Company, 1966.

Post, T. J., & Price, H. E. Requirements and criteria for improving reading comprehension of technical manuals. BioTechnology, Inc. for Naval Sea Systems Command, Technical Publications Branch, Contract N00024-74-C-5426, November 1974.

Rathbone, R. R., & Stone, J. B. A writer's guide for engineers and scientists. New Jersey: Prentice-Hall, Inc., 1962.

Smith, R. W. Technical writing. New York: Barnes & Noble, Inc., 1963.

Strunk, W., Jr. The elements of style (2nd ed.). New York: The Macmillan Company, 1972.

Walton, T. F. Technical manual writing and administration. New York: McGraw-Hill Book Company, 1968.

Wood, G. M. Suggestions to authors of papers submitted for publication by the United States geological survey. Washington, D. C.: United States Government Printing Office, 1935.

APPENDIX

TECHNICAL DOCUMENTATION PROCESSING SYSTEMS AND DEVICES

TECHNICAL DOCUMENTATION PROCESSING SYSTEMS AND DEVICES

This appendix provides information in detail on selected systems and devices in existence that are capable of performing one or more of the steps involved in the processing of text and graphics for technical documentation. It is intended as a representative sample of such systems and devices. It is readily acknowledged by the author that systems of equal or greater capabilities no doubt exist but did not come to his attention in this investigation.

1. American Telephone and Telegraph "TOPP" System.

AT&T has made great strides in solving essentially the same problems that the Navy is encountering regarding technical documentation.[1] Bell Laboratories at Whippany, New Jersey, began work in the early 1960s on the development of a Maintenance Data System (MDS) as part of the SAFEGUARD antiballistic missile (ABM) program. The Army accepted MDS in 1965 as their ABM maintenance approach. It not only increased technician productivity by 20 to 30 percent but also resulted in lower-skill technicians performing routine and trouble-shooting tasks almost as effectively as highly skilled individual. The Task-Oriented Practices - Plant (TOPP) was developed out of this approach for internal AT&T use.

The preliminary findings on the prototype system were particularly notable for the low-service technician:

a. Productivity increased by 34 percent.

b. Errors were reduced by 50 percent.

c. Task completion increased by 114 percent. (AT&T, 1974, p. 1)

It was therefore concluded that the prototype made the short service employee almost as effective as his much longer service counterpart. On the basis of these results, AT&T Plant and Engineering, Bell Telephone Laboratories, and Western Electric Technical Publications jointly developed a standard that would define the methods and criteria for applying the new approach to selected system equipment. This standard, in effect, became a TOPP on how to prepare a TOPP and provides detailed step-by-step guidance to the TOPP writer. In essence, this appears to be much the same (conceptually) as a Military Specification given to a writer of Navy manuals. TOPP materials developed by writers using this standard (for three items of equipment) were evaluated in five operating companies with the following result: productivity was improved by 19 percent for a low-skill group, 36 percent for an average-skill group, and 16 percent for a high-skill group. (No information on error rate or task completion was given for this evaluation.) The author has examined a sample TOPP and found it to be exceptionally well designed and suitable for variable levels of personnel capabilities. In addition to the above data, users were asked to rate TOPP and traditional BSPs ("Bell System

---

[1]This information (including all quotations) is found, for the most part, in an AT&T internal document titled "A New Approach to Maintenance BSPs: Task-Oriented Practices - Plant (TOPP)," January 1974. Permission to reference this source was kindly provided by AT&T.

Practices," the forerunner of TOPP) on a scale from 1 (a perfect document) to 5 (a worthless document). The mean rating for TOPP (across the three skill levels) was 1.8 as opposed to 4.8 for traditional BSPs. A high degree of user "acceptability" was therefore indicated. On a similar scaling basis, monitor judgment of employee performance averaged 2.2 (with 1.0 optimal) for TOPP and 4.7 for BSPs.

In summary, the following advantages seem inherent in the TOPP system:

a. TOPP is based on a systems approach to the maintenance task, including thorough and extensive task analysis.

b. The system is supported by a set of "standards" to ensure that development of a high-quality product, from task analysis to evaluation of the entire system.

c. TOPP considers in depth the "Training - TOPP trade-off" (a concept analogous to that of the "Head - Book trade-off" within DoD).

d. The system uses a flow chart method for detailed level procedures rather than text (making it more effective for low-skill personnel--those who use the detailed procedure most often).

e. TOPP addresses troubleshooting in addition to other maintenance tasks--an area generally causing a great deal of difficulty in traditional technical manual writing.

f. The system is compatible with book, microfilm, computer store (with CRT access), audiovisual devices, etc.

g. Individual TOPPs contain not one, but several, cross-indexing methods for rapid access from one section to another.

h. To the greatest possible extent, TOPP is bound according to tasks (similar to the Navy's "work package" concept), such that minimal cross-referencing to separate volumes is required.

i. As a part of the TOPP system, graphics (schematics, flow charts, etc.) are designed at an interactive computer terminal, requiring only a rough free-hand sketch by the engineer. The designer is, in general, a clerical person with some artistic talent. Once in "best form" at the console, the computer outputs a microfilm image of the graphic from which a direct final copy print is made.

Note the congruence of this overall system with the "themes" expressed at the earlier-discussed conference on JPAs. The system incorporates many of the concepts which, by consensus, are necessary in a Navy counterpart (e.g., task analysis, multiple level of detail, etc.). And more notably, TOPP is a system which is constructed with a model in mind. It could, perhaps, become a prototype for a Navy manual production model.

2. The Navy TRUMP/MIARS System.

It has been emphasized that we should use today's technology to the utmost to solve our immediate problems and rely on tomorrow's research to counter long-range problems. The Navy has, on a test and evaluation basis, a system which incorporates much of the best of today's capabilities: Technical Review and Update of Manuals and Publications (TRUMP) and the Maintenance Information Automated Retrieval System (MIARS). TRUMP is intended ultimately as a total system for the revision and update of technical manuals. MIARS is an adjunct to TRUMP and is basically a microfilm storage and retrieval system with associated reader/printers. As it is not an integral part of the technical manual production process, it will not be discussed further here.

TRUMP is capable of converting documents from their standard hard copy form (via microfilm) to digital form by means of Optical Character Recognition (OCR), thus eliminating the need for keypunching to card or tape. As reported by the Naval Air Systems Command, "A typical TRUMP configuration . . . has input capacity challenging the rate of 40 to 50 input keyboard operators."[2] Once in the data bank, the manuscript is available for selective recall for the purposes of alternations, updating, delection, etc., and subsequent reproduction in hard copy form via computer output microfilm. Illustrations and tables are not presently digitized, but rather are photographed and stored to be recalled when needed for photocomposition.

Aside from OCR, a major feature of TRUMP is an interactive computer capability for use in editing, proofreading, and page composition. This feature is basically a cathode ray tube on which a page or portion of a page is presented, a keyboard by which entries can be made to the selected page, and a "data tablet" on which a stylus can be used to position data on a page and manipulate that data in various ways. The CRT and keyboard can also be readily used for "reject correction" (i.e., when a symbol cannot be accurately read by the optical scanner).

TRUMP, designed in large part by Information International, Incorporated, is currently undergoing test and evaluation at the U.S. Naval Air Rework Facility, Jacksonville, Florida. In addition to the capabilities discussed above, it would seem to have the potential for much more sophisticated features. Readability analysis is not currently a part of the TRUMP repertoire, but could quite easily be incorporated. Writing aids, such as a stored lexicon or a key-word-in-context list could be incorporated and used in conjunction with the interactive terminal. The OCR capability of the system, together with its microfilm reading characteristics, could well be a large step toward the solution of problems of filing, storing, printing, and updating of data coming from sources quite remote from the system itself (microfilm being much easier to transfer from one site to another than hard copy). As an example, 22 sets of manuals, each consisting of 101 volumes, were replaced by 5 reader/printer machines and microfilm on an A-4 aircraft assembly line. The microfilm for each set, replacing approximately 48,000 pages of hard copy, was contained in only 16 cartridges (Naval Air Systems Command, 1974).

---

[2]Technical Data Management Projects of the Naval Air Systems Command publication "MIARS and TRUMP Mechanized Update Techniques," October 1974.

3. Time-Sharing System.[3]

The time-sharing system described here is basically a multipurpose computer tool used in such diverse areas as information management, financial planning, inventory management, and others. Its applicability to Navy technical writing lies with its capability to provide timely information on the readability of written material and its potential for future sophistication of the analysis process. The readability program (basically a Flesch Reading Ease analysis) can be used by any keyboard operator working interactively with the computer time-sharing program. Text is entered on a terminal, and immediate analysis of the text is received. Some of the features and implications for future use are as follows:

a. It can be used on a nation-wide basis--an advantage for Navy technical writing because of the many geographically separated sites where such writing is performed.

b. It would be cost-effective for computerized readability analysis because a core computer would not be required at every writing site. It is reported that terminal costs can range from about $80.00 per month to about $250.00 per month depending on the type of unit used.

c. It has a feature which might help overcome the problem of technical terminology "inflating" the difficulty level of writing. In the typing process, words which are not to be included in the readability analysis can be put in brackets. These words are counted as words, but not as difficult words.

d. The system has a voluminous storage capacity, so a manual could be stored in its entirety for some period of time (e.g., up through the verification process) and then instantaneously recalled for updating or correcting.

e. The system is accompanied by a relatively simple programming procedure by which programs can be written with very little training to perform different analyses on writing, such as "calling out" the locations of any word or phrase the writer might desire.

f. It has the capability for any site on the system to communicate almost instantly with any other site, which may have advantages for interchange of information between contractors and contract monitors (for example).

In summary, a system with the above capabilities and potential could be a very useful tool for the improvement of Navy instructional and technical manual production.

---

[3]The system upon which this discussion focuses is that of Scientific Time Sharing, Inc. It is the system itself, and not its marketer that is being featured. No bias for or against Scientific Time Sharing, Inc. is intended.

4. <u>On-Line Pattern Analysis and Recognition System (OLPARS)</u>.

OLPARS is a high-level system, the use of which requires little or no programming skill or knowledge of the details of the mathematics used. It presents physical objects or events on a Cathode Ray Tube (CRT) at which an operator with a light pen performs operations on the presentation. The system may be used in this way to determine correlations between properties of the object or event, to determine correlations between values of these objects, and to generate, test, refine, or reject different property classification schemes. All use of the system is on-line to provide fast response and user interaction. It can be used for pattern analysis and pattern recognition (among other uses), both of which have implications for measuring the usability of graphics, refinement of graphics on-line, and the potential for microfilm or hard copy output of the final product. With such an output capability, the system could be used to systematically vary dimensions or values of a particular graphic, obtain hard copy of each, and present them to qualified judges for determination of their relative comprehensibility or usability. Because of the versatility of such a system, it is likely that other valuable research and/or production ideas could be generated. This system is currently undergoing test and evaluation at the Naval Ship Research and Development Center, Carderock, Maryland.

5. <u>Readability Ease Assessment Device (READ)</u>.[4]

This device involves an electronic tabulator/calculator connected to a modified electric typewriter. The tabulator counts strokes as material is being typed to determine letters per word and words per sentence. READ then calculates the solution to a readability regression equation (derived from Smith & Senter, 1967) and the Reading Grade Level of the material is immediately displayed to the typist in digital form. While certainly not as flexible as a computerized system, READ does offer the writer information about the reading complexity of his material while it is still in draft version. In this way, writers may edit the material in their normal fashion and make changes for improved readability if required. Its major advantage, then, is that it offers real-time feedback on the readability of written materials at an early stage in the production process.

6. <u>Automated Data Presentation Electronic Photo Composition (ADPEP)</u>.

This system was designed with the following objectives in mind (among others): to have high reliability, to be easy to operate, to be cost-effective, and to perform all formatting functions--text, double galley, full page, and tabular material. It is basically a small computer with Selectric typewriters used for input onto magnetic tape. A line printer output is then provided for proofing and correction. The text (but not the illustrations) are then composed by computer and provided in hard copy form into which illustrations are merged. The final copy is then photographed for printing. This system, while offering potential for fully automated production of Navy manuals, has some limitations which the Navy's TRUMP system (for example) does not. It requires typed input,

---

[4]This device, and the ADPEP system to be described below, were developed by McDonnel Douglas Corporation. No bias for or against this manufacturer is intended.

rather than using Optical Character Recognition (OCR). It is not at present capable of merging illustrations with text automatically. It is offered here simply as one other example of existing systems having potential for the production of manuals.

7. Computer Aided Revising, Editing, and Translating System (CARET I).

After typing, in any stage of the draft process, this system produces a computer printout which has several characteristics to make editing or revising easier. The printout is triple-spaced to permit changes to be made directly. The number of syllables for each word is printed under that word. The number of words in each sentence is printed under the first word of that sentence. Each sentence begins at the left margin of the printout so that sentence pattern can be easily seen and unusually long sentences noted. At the end of each sample of text, readability formula scores are provided based on any one or all of five different formulas. The system was developed by G. R. Klare and associates at Ohio University and Harvard University (Klare, Rowe, St. John, & Stoulurow, 1969).

Beyond the basic editing function, the system also has the capability for automatically printing out "cloze tests" for examining the comprehensibility of material. Potentially, the system offers many valuable advances over its present capacity. One example, proposed for the future, is the capability of constructing a dictionary or lexicon of words contained in the input text and from this building a "key-word-in-context" (KWIC) capacity. This latter function will take any word selected by the writer or editor and print out every instance of that word in the text. The format generally follows the rule that instances of the selected word appear in a column down the middle of a page and a certain number of words which surround each key word in the text are printed out to its right and left. Such a function then allows the writer and/or editor to make possible changes in sentence construction to enhance the readability and comprehensibility of the material. This capability, particularly if on-line, could greatly enhance the speed and flexibility of rewriting for better readability. The planning for this system has been quite thorough and systematic; forward steps appear to have been taken only when current ones were reliabily operational. It is considered to have the potential for virtually all the capabilities discussed in connection with the processing and production of Navy training and technical manuals.

DISTRIBUTION LIST

Chief of Naval Operations (OP-914), (OP-964D), (OP-987H), (OP-401), (OP-99), (OP-991B)
Chief of Naval Personnel (Pers-10c), (Pers-2B)
Chief of Naval Material (NMAT 03424), (NMAT 0422)
Chief of Naval Research (Code 450) (4)
Chief of Naval Education and Training (00A), (003), (N-2), (N-4), (N-5)
Chief of Naval Technical Training
Chief of Naval Technical Training (016), (N-3), (N-4)
Chief of Naval Education and Training Support
Chief of Naval Education and Training Support (01A)
Chief of Information (OI-2252)
Director of Navy Laboratories
Commander Training Command, U.S. Atlantic Fleet (Code N3A)
Commanding Officer, Fleet Combat Training Center, Pacific (Code 00E)
Commanding Officer, Naval Education and Training Program Development Center
Commanding Officer, Naval Development and Training Center (Code 0120)
Commanding Officer, Naval Aerospace Medical Institute (Library Code 12) (2)
Commanding Officer, Naval Training Equipment Center
Commanding Officer, Naval Air Development Center
Commanding Officer, Naval Amphibious School, Coronado
Commanding Officer, Naval Damage Control Training Center
Commanding Officer, Fleet Training Center, Naval Station, San Diego
Commander, David W. Taylor Naval Ship Research and Development Center (186) (6)
Commander, Naval Air Systems Command (04A4), (340C)
Officer in Charge, Naval Education and Training Information Systems Activity, Memphis Detachment
Director, Training Analysis and Evaluation Group (TAEG)
Director, Navy Occupational Task Analysis Program Department, Naval Personnel Program Support Activity
Military Assistant for Training and Personnel Technology, ODDR&E, OAD(E&LS)
Director for Acquisition Planning, OASD(MRA&L)
Personnel Research Division, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
Occupational and Manpower Research Division, Air Force Human Resources Laboratory, Brooks Air Force Base
Technical Library, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base
Flying Training Division, Air Force Human Resources Laboratory, Williams Air Force Base
Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base
Program Manager, Life Sciences Directorate, Air Force Office of Scientific Research (AFSC)
Army Research Institute for the Behavioral and Social Sciences
Coast Guard Headquarters (G-P-1/62)
Defense Documentation Center (12)